Universidade de Lisboa

Faculdade de Farmácia

Application and Verification of Next-Generation Sequencing (NGS) for the Molecular Diagnostics of Brain Tumours

Maria de Azevedo Pinto

Mestrado Integrado em Ciências Farmacêuticas

2017

Universidade de Lisboa

Faculdade de Farmácia

Application and Verification of Next-Generation Sequencing (NGS) for the Molecular Diagnostics of Brain Tumours

Maria de Azevedo Pinto

Trabalho de Campo de Mestrado Integrado em Ciências Farmacêuticas apresentada à Universidade de Lisboa através da Faculdade de Farmácia

Orientador: Doutor Andreas Waha, Professor Assistente
Coorientador: Doutor Rui Eduardo Castro, Professor Assistente

2017

universitätbonn

Erasmus+

The work that led to the production of this dissertation was carried at the Hirntumor-Referenzzentrum of the Institut für Neuropathologie of the Universität Bonn, under Erasmus+ Programme.

Supervisor: PD Dr. Andreas Waha

To my father

Abstract

Primary brain tumours are a critical cause of morbidity and mortality in both adults and children, representing around 1-2% of all newly diagnosed tumours and accounting for about 2% of all cancer-related deaths. Gliomas are the most prevalent primary malignant brain tumour representing 80% of these. Over the past years, distinctive genetic profiles have been identified in several glioma types which have led to the WHO 2016 new classification of CNS tumours that incorporates molecular parameters into the tumour classification criteria, breaking with the previous approach based entirely on histological features. This refines tumour diagnostics and can also provide important prognostic and therapeutic response information. The aim of this study was to apply and verify if Next-Generation Sequencing (NGS) can effectively replace the classical methods – pyrosequencing and Sanger sequencing - in establishing the molecular diagnostics of gliomas. Thus, a glioma-tailored gene panel covering 518 amplicons of 19 genes frequently aberrant in gliomas was designed and applied to assess 30 glioma samples. This targeted NGS approach was carried out by Illumina® TruSeq® technology in Illumina® MiSeq System. DNA libraries preparation showed a success rate of 90%. Data analysis was performed using several bioinformatical software and filtering parameters of variants were optimized to reduce sequencing artefacts in the NGS run. Better DNA quality libraries presented more reliable results showing less DNA sequence changes. The sensitivity and specificity of the 19-gene panel for detection of DNA sequence variants were verified by single-gene analyses which showed to be substantial concerning hotspot mutations but not so trustworthy concerning new-mutations detected by NGS. The presented findings showed that, even though NGS application in routine glioma molecular diagnostics can't be yet implemented, further investigation of this technology is promising since NGS showed to be a resourceful tool for glioma genetic profiling, displaying its potential as diagnostic method which would facilitate the integrated histological and molecular glioma classification.

Key-words: brain tumours, glioma, molecular diagnostics, mutation, Next-Generation Sequencing (NGS)

Resumo

Os tumores cerebrais primários são uma causa crítica de morbilidade e mortalidade em adultos e crianças, representando cerca de 1-2% de todos os tumores recém-diagnosticados e cerca de 2% de todas as mortes relacionadas com o cancro. Os gliomas são o tipo de tumor cerebral maligno primário mais prevalente representando 80% destes. Ao longo dos últimos anos, foram identificados perfis genéticos distintivos em vários tipos de glioma o que levou à nova classificação de tumores do SNC de 2016, pela OMS, que incorpora os parâmetros moleculares nos critérios de classificação do tumor, quebrando com a abordagem anterior inteiramente baseada nas características histológicas. Esta nova abordagem veio aperfeiçoar o diagnóstico dos tumores e a capacidade de fornecer, também, informações importantes quanto ao prognóstico e resposta à terapêutica. O objetivo deste estudo foi aplicar e verificar se a Sequenciação de Nova Geração (NGS) pode efetivamente substituir os métodos clássicos - pirosequenciação e sequenciação de Sanger - no estabelecimento do diagnóstico molecular dos gliomas. Assim, foi desenhado e aplicado um painel genético adaptado a gliomas que cobriu 518 amplicões de 19 genes frequentemente aberrantes nestes, para analisar 30 amostras de gliomas. Esta abordagem direcionada da NGS foi realizada pela tecnologia TruSeq da Illumina®, no seu sistema MiSeq. A preparação das bibliotecas de DNA mostrou uma taxa de sucesso de 90%. A análise dos dados foi realizada recorrendo a vários softwares bioinformáticos e os parâmetros de filtração das variantes obtidas foram otimizados para reduzir os artefactos da sequenciação resultantes da execução da NGS. As bibliotecas de DNA de melhor qualidade apresentaram resultados mais confiáveis exibindo menos alterações nas sequências de DNA. A sensibilidade e a especificidade do painel de 19 genes para a deteção de mutações foram verificadas por análise individual dos genes, indicando ser substanciais em relação às mutações hotspot, mas não tão confiáveis em relação às novas mutações detetadas pela NGS. Os resultados apresentados mostraram que, apesar de não ser possível implementar para já a NGS na rotina do diagnóstico molecular de gliomas, é promissora uma investigação adicional nesta tecnologia, uma vez que a NGS mostrou ser uma ferramenta rica em recursos para delinear o perfil genético dos gliomas, ilustrando o seu potencial como método de diagnóstico que facilitaria a classificação histológica e molecular integrada dos gliomas.

Palavras-chave: tumores cerebrais, glioma, diagnóstico molecular, mutação, Sequenciação de Nova Geração (NGS)

Acknowledgements

I would like to grant my deepest thanks to Prof. Dr. Andreas Waha for giving me the opportunity of working on this research project, at the Hirntumor-Referenzzentrum of Germany. I am deeply grateful for your trust, advice and constant motivation and companionship. It was an honor to form the "team NGS" with such a great scientist and person and, above all, it was an honor to have your guidance.

To Prof. Dr. Rui Eduardo Castro for the fundamental final counselling and support. I am very thankful for your prompt availability, help and guidance.

I want to particularly thank to Verena Dreschmann and Evelyn Doerner for the technical support and advice. I am very grateful for your patience, availability and hospitality.

I am also very thankful to Prof. Dr. Bernd Evert for the tireless motivation and the entertaining "knowledge exchange" sessions in the pauses of the lab work.

To my mother, Marta Pinto, for your love and ability for encouraging me in your singular way. I am deeply grateful for having you always by my side.

I am profoundly grateful to my uncles, Maria Inês Iglésias and João Tiago Iglésias, for all financial and emotional support. I am very thankful for your constant encouragement and help.

I am very thankful to my grandparents, Maria Fernanda Mendoça e José Eduardo Mendoça, for the ongoing support. Specially to my grandmother, thank you for always having the right word at the right time.

To my cousin, Mariana Freitas, for being always present, especially when I need it most. I am very grateful for your comforting understanding and constant motivation.

To my brother, António Pinto, for the backing and incessant optimism.

Finally, I would like to thank to all my friends who supported me during this important phase of my life, particularly to: Sandra Silva, David Pereira, Mariana Alexandre, Alexandra Ângelo, Leonardo Lourenço, André Marques, Jéssica Bronze, Márcio Santos, Jorge Honda, Joana Fernandes, Milene Fortunato, Pedro Maçãs, Cibele Maçãs, Renato Guerreiro, João Matias and Ruben Mendes – because… If I could do this without you? I could, but it wouldn't be the same!

Abbreviations

ºC	degree Celsius
µg	microgram
µl	microliter
5-mC	5-methylcytosine
2800M	control DNA 2800M
A	adenine
A5XX	index i5 adapters
A7XX	index i7 adapters
ACD1	amplicon control DNA
ACP3	amplicon control oligo pool 3
ACVR1	activin A receptor type 1
APS	adenosine 5’ phosphosulfate
ATP	adenosine triphosphate
ATRX	alpha thalassemia/mental retardation syndrome X-linked
bp	base pair
BR	broad range
BRAF	B-Raf proto-oncogene, serine/threonine kinase
C	cytosine
CAT	custom amplicon oligo tube
CDK4	cyclin-dependent kinase 4
CDK6	cyclin-dependent kinase 6
CDKN2B	cyclin-dependent kinase inhibitor 2B
Chr	chromosome
CIC	capicua transcriptional repressor
CLP	clean-up plate
CNS	central nervous system
COSMIC	catalogue of somatic mutations in cancer
DAL	diluted amplicon library
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate

ddNTP	dideoxyribonucleotide triphosphate
dGTP	deoxyguanosine triphosphate
DIPG	diffuse intrinsic pontine glioma
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dsDNA	double-stranded DNA
dTTP	deoxythymidine triphosphate
e.g.	for example
EDP	enhanced DNA polymerase
EDTA	ethylenediaminetetraacetic acid
EGFR	epidermal growth factor receptor
ELB	extension-ligation buffer
ELE	extension-ligation enzyme
EMM	enhanced master mix
ERBB2	Erb-B2 receptor tyrosine kinase 2
EtOH	ethanol
FFPE	formalin-fixed, paraffin-embedded
FGFR1	fibroblast growth factor receptor 1
FISH	fluorescence in situ hybridisation
FUBP1	far upstream element binding protein 1
fw	forward primer
g	gram
G	guanine
Gb	gigabase pairs
GBM	glioblastoma multiforme
GcGBM	giant-cell glioblastoma multiforme
gDNA	genomic DNA
h	hour
H3F3A	H3 histone family member 3A
HGG	high-grade glioma
HIST1H3B	H3 histone cluster 1 family member B
HPLC	high-performance liquid chromatography
HT1	hybridization buffer

HYP	hybridization plate
IARC	international agency for research on cancer
IDH1/2	isocitrate dehydrogenase 1/2
IGV	integrative genomics viewer
l	liter
LGG	low-grade glioma
LNA1	library normalization additives 1
LNB1	library normalization beads 1
LNP	library normalization plate
LNS2	library normalization storage buffer 2
LNW1	library normalization wash 1
LOH	loss of heterozygosity
m^2	square meters
M	molar
mg	milligram
MGMT	O(6)-methylguanine-DNA methyltransferase
min	minute
ml	milliliter
MTP	microtiter plate
n	number
N	normal
NaOH	sodium hydroxide
NF1	neurofibromin 1
ng	nanogram
NGS	next-generation sequencing
NOS	not otherwise specified
NTC	no template control
OHS2	oligo hybridization for sequencing reagent 2
PAL	pooled amplicon library
PCR	polymerase chain reaction
PDGFRA	platelet-derived growth factor receptor A
PF	passing filter/reads identified
PIK3CA	phosphoinositide-3-kinase catalytic subunit alpha

PIK3R1	phosphoinositide-3-kinase regulatory subunit 1
PPi	pyrophosphate
PPM1D	protein phosphatase, Mg^{2+}/Mn^{2+} dependent 1D
ps	pyrosequencing primer
PTEN	phosphatase and tensin homolog
RELA	NF-Kappa-B Transcription Factor P65
rev	reverse primer
ROI	region of interest
rpm	rotations per minute
RS1	resuspension solution 1
RSB	resuspension buffer
SAV	Illumina sequencing analysis viewer
sec	second
SGP	storage plate
SNV	single-nucleotide variation
SPB	sample purification beads
SS1	sample stabilization solution 1
SW1	stringent wash 1
T	thymine
TBE	tris-borate-EDTA
TERT	telomerase reverse transcriptase
TMZ	temozolomide
TP53	tumour protein P53
Tris-HCl	tris(hydroxymethyl)aminomethane-hydrochloride
U	uracil
UCSC	university of california santa cruz
UV	ultraviolet
V	volt
WHO	world health organization
WT	wild-type

Table of Contents

Figures Index

Tables Index

1 INTRODUCTION

"Tumours of the central nervous system (CNS) represent a relatively rare but serious health burden." (1)

In Europe, the estimated age-adjusted annual incidence of tumours of the Central Nervous System in 2012 was 6,6 cases per 100.000 individuals and its mortality rounded the 4,9 cases per 100.000 individuals (Figure 1), according to the International Agency for Research on Cancer (IARC) (2). In that year, 57.000 new diagnoses and 45.000 deaths regarding tumours of the brain and other parts of the CNS were estimated in Europe (3).

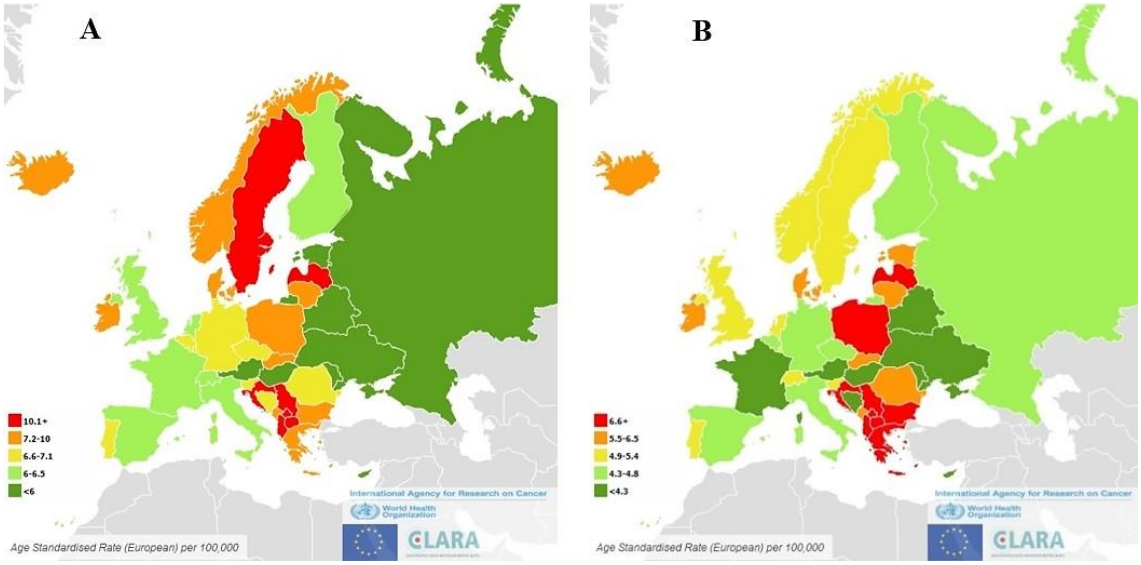


Figure 1 - Estimated incidence (A) and mortality (B) of brain and CNS cancer in both sexes in Europe, 2012
Illustrative representation of the number of individuals who were diagnosed, **A**, and died, **B**, for brain or other CNS tumour in 2012, in European countries; Western and Eastern countries are less affected when compared to central Europe; despite having high incidence, Northern countries own the lowest number of mortality reports associated with brain and other CNS cancer. (figure adapted from (2))

Brain tumours are classified into primary and secondary. Primary brain tumours are originated in the brain itself and they can be either malignant or benign oppositely to secondary brain tumours, which are metastasis originating from a tumour located in another part of the body and they are always malignant. These intracranial neoplasms are also subdivided into extracerebral tumours - meningiomas -, and intracerebral tumours which can be intrinsic – gliomas -, if they arise from the parenchyma of the brain or extrinsic - lymphoma or metastasis (4).

Primary brain tumours are a critical cause of morbidity and mortality in both adults and children (5), representing around 1-2% of all newly diagnosed tumours and accounting for about 2% of all cancer-related deaths (4,6). These tumours regularly cause severe disabilities and generate great distress not just in the patients, but also in their families and in the health care systems (5).

Gliomas represent approximately 30% of all CNS tumours (7) and they are the most prevalent primary malignant brain tumours being responsible for 80% of these last mentioned (8,9,10). Worldwide, the annual incidence of these primary intracerebral neoplasms is about 6 cases per 100,000 individuals (11) and it is known that the western world has higher incidence of gliomas comparing to the less developed countries which could be explained by the deficient reporting of gliomas cases, limited access to health care and discrepancies in diagnostic practices (10). Glioma is derived from the Greek word for "glue", as they emerge from CNS cells which form the scaffolding and structural support for neurons – astrocytes or oligodendrocytes (12). So, as gliomas occur in the glial tissue of the CNS, they share characteristics of the normal glial cell type that they most closely resemble histologically (7). Traditionally, these tumours are divided into two major categories based on the degree of invasiveness of the brain parenchyma: (i) gliomas with "circumscribed" growth behaviour and (ii) diffuse gliomas with diffuse infiltration of the brain parenchyma (13). Thus, gliomas are classified in line with those histological similarities and graded from I to IV based on morphology and malignant behaviour, according to the standard for nomenclature and diagnostics of CNS tumours: the World Health Organization (WHO) classification and grading system (8,3,14,10). The most typical gliomas found in adults include glioblastoma (grade IV), astrocytomas (grade I–III), oligodendrogliomas (grade II–III) and ependymomas (grade I–III) (8,14). It is known that there is a strong correlation between grade and prognosis: the higher the grade, the poorest the prognosis (3).
Despite having diverse patterns of genomic variations and clinical outcomes, glioma patients most frequently bear a fatal prognosis which entails a substantial impact on them and their families' physical, psychological and social status (8,15,16,17).

The recognition of important genetic, epigenetic and transcriptional abnormalities in the several types of gliomas revolutionized the understanding of its physiopathology, its pathological classification, the patient treatment and novel therapeutic approaches, admitting that particular molecular alterations serve as diagnostic markers for more

accurate classification and others are associated with therapy response and prognosis (17,18).

The aim of this study was to examine the success and the advantages of using a modern parallel DNA sequencing technology – the Next-Generation Sequencing (NGS) – over the conventional single-gene-approach technologies – pyrosequencing and Sanger sequencing - in performing the molecular diagnostics of gliomas, mainly of previously histologically classified WHO grade IV glioblastomas, the most aggressive, invasive and undifferentiated type of brain tumour (10).

1.1 Glioblastoma multiforme (GBM)

WHO grade IV glioblastoma multiforme (GBM) is the most aggressive, invasive, undifferentiated and frequently occurring malignant type of brain tumour, accounting for more than 60% of all brain tumours (6,7,10). This malignancy is distinguished by its high degree of cellularity, microvascular proliferation, tumour cell chemoresistance and necrosis (6).

GBMs are subdivided according to their clinical history as either primary GBM or secondary GBM, evolving through distinct genetic pathways, affecting patients at different ages and having various outcomes. Primary GBMs (arising *de novo,* in the absence of any precursor lesion), frequently *IDH*-wildtype, account for approximately 90% of GBMs and is seen in older patients (mean age 62 years) with a rapid and robust disease progression. On the other hand, secondary GBM (arising from malignant progression of lower-grade astrocytoma or oligodendroglioma), generally *IDH*-mutant, occur in younger patients (mean age 44 years) with a longer history of disease progression and a better prognosis than primary GBMs (19,9,20,7,14).

Despite the diversity of modern therapies, GBM is still an incurable disease with terrifically poor prognosis and a median survival of 15 months (19,10).

1.1.1 Epidemiology

GBM is considered the most prevailing brain and CNS malignancy in adults, representing approximately 50% of all gliomas (85% of malignant gliomas), 60% of all brain tumours (80% of malignant brain tumours) and 17% of all primary brain and CNS neoplasms (45,2% of malignant primary brain and CNS neoplasms) (10,19). About 17.000 new cases of this grade IV glioma are diagnosed per year (21).

GBM can occur at any age but it is mainly diagnosed at older ages, with the median age of diagnostics of 64 years, and its incidence increases with age, being rare in paediatric population, uncommon before 20 years of age and reaching the peak at the 7^{th}-8^{th} decade of life (8,9,19,10).

Men have a higher incidence of GBM as compared to women (information ratio approximately 1,5 times greater); however, secondary GBMs seem to have the same or

even higher frequency in women. It is also known that white individuals have the highest rates for GBM when compared to the other ethnic groups. (19,10,21)

1.1.2 Etiology and risk factors

There is limited knowledge about the etiology of these highly incurable brain tumours (10). Exposure to therapeutic or high-dose ionizing radiation is the only strongly established direct cause of GBM even after studying many genetic and environmental factors (9,22,10,19). Thus, GBM is considered mostly sporadic just like many other cancers (9,19).

GBMs are reported to occur in families but there is no susceptibility gene identified yet. However, an association between some rare genetic disorders and an increased GBM incidence is established. So, screening is only performed in these individuals at genetic risk – e.g. neurofibromatosis type 1 and 2, tuberous sclerosis, Li-Fraumeni syndrome and Turcot syndrome (22,10,9,11). Also, several single-nucleotide polymorphisms (e.g. *CDKN2B* on chromosome 9p21) seem to be linked to an increased risk of GBM. These variants, found in at least 1% of a population, might disturb the physiological function of the gene product, accounting for individual differences in disease risk in sequence with other variants or exposures (19).

Allergies or atopic diseases such as psoriasis, asthma and eczema, are associated with lower risk of GBM probably due to the activation of immune surveillance mechanism having a protective effect (23,24). Anti-inflammatory therapy use for short periods of time (<10 years) is also correlated with protective effect against GBM, particularly when there is no clinical history of allergies (25).

GBM was linked to high-dose chemotherapy for cancer treatment at loci other than the brain, but the entity which determine the degree of risk from the exposure is individual genetics (22).

To date no other significant evidence exists associating GBM with environmental factors or life-style aspects such as dietary exposure to N-nitroso compounds (e.g. cured or smoked meat or fish), smoking, alcohol consumption and drugs usage, pesticide exposure, severe head injury or even exposure to diagnostic radiation. Additionally, no conclusive association has been found between the use of mobile phones and the risk of GBM (19,22,26,27).

1.1.3 Prognosis, treatment and survival

In clinical oncology, GBM treatment remains one of the most challenging assignments, despite all international attempts. Over the past few years, diverse therapies have been investigated with very limited success (19).

The location and the complex heterogeneous nature of these tumours is the great responsible for the difficulty in effectively treating GBM patients (28,19).

Presently, standard management of GBM includes not only specific anti-tumour therapy but also implementing efficient supportive care to the patient in order to handle the various symptoms and signs of the disease (e.g. corticosteroids for neurological symptoms relief; Levetiracetam for patients with seizures). The current accepted antineoplastic treatment for GBM patients involves neurosurgical resection followed by radiation therapy and concurrent aggressive chemotherapy with oral temozolomide (TMZ) 75 mg/m² daily and then succeeded by six cycles of maintenance treatment with oral TMZ 150–200 mg/m² for 5 days every 4 weeks. However, in most cases recurrences emerge and they can be managed with additional surgery and adjuvant treatment with antiangiogenic (e.g. Bevacizumab) or chemotherapeutic (e.g. Carmustine and Lomustine) drugs. Marginal benefits are observed but patients most frequently die (7,28,19,11,29).

Hence, with no effective treatment against GBM in the current time, this malignancy remains incurable carrying a deadly poor prognosis. GBM patients have a median life-expectancy of only 15 months after diagnostics and the 5-year survival is 2-4% (21,6,30).

1.1.4 Site and clinical presentation

GBM is most frequently situated in cerebral hemispheres, occurring essentially in the supratentorial region (primary GBMs arise in any lobe with a widespread distribution, whereas secondary GBMs have a striking predilection for the frontal lobe) and only few percent of these neoplasms arise in cerebellum (0.4-3.4%) and even more rarely in the brainstem and spinal cord (31,19).

GBM patients often present a short clinical history, ranging between 1 to 6 months. For patients with secondary GBM clinical history can extend to a few years (10,8,32). Different signs and symptoms may be present in GBM patients, some are results of the tumour direct effect and others of secondary effects of increased intracranial pressure

(10). Hereby, GBM clinical presentation features include essentially focal neural deficits - the most common symptom (40-60%) -, neurocognitive impairment, epileptic seizures, headache, vomiting and altered consciousness (11,10,8).

1.1.5 Macroscopic and histological features

Macroscopically, the tumour is typically represented by single, relatively large, irregular shaped lesion having a ring-shaped zone of contrast enhancement around a dark, central area of necrosis, which usually arises in the white matter (Figure 2). Regardless of their feature heterogeneity, GBM is known for significant and consistent multifocal haemorrhage, necrosis and exuberant vascular hyperplasia (30,33,34).

In addition to imaging findings a tissue analysis is required for definitive histopathological confirmation and to discriminate from other primary and metastatic brain tumours (35).

Histologically, GBM presents hypercellularity with a pleomorphic cell population varying from small poorly differentiated tumour cells to large multinucleated cells, nuclear atypia, prevalent mitotic activity and its major characteristic features of microvascular proliferation, frequently with glomeruloid structure and multifocal necrosis with pseudopalisading cells (Figure 3) (13,16,33,10,36).

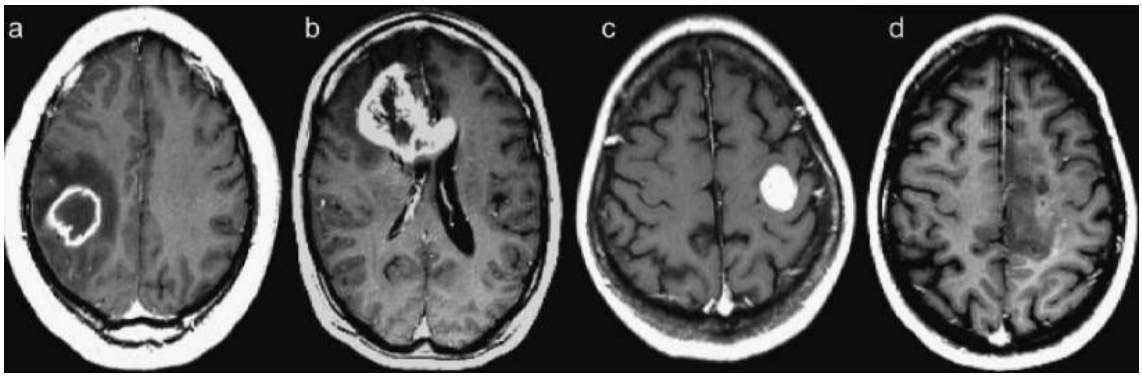


Figure 2 - Representative radiological findings of GBM in four different patients
Contrast-enhanced axial images illustrate the heterogeneity of the tumour: (a) rim-enhancing mass with central necrosis and surrounding edema in the right parietal lobe; (b) irregularly enhancing mass crossing the corpus callosum; (c) well-circumscribed homogeneously enhancing mass with no combined edema in the left frontal lobe; (d) poorly-defined infiltrative mass in the left medial frontal lobe with no significant necrosis. (figure origin: (33))

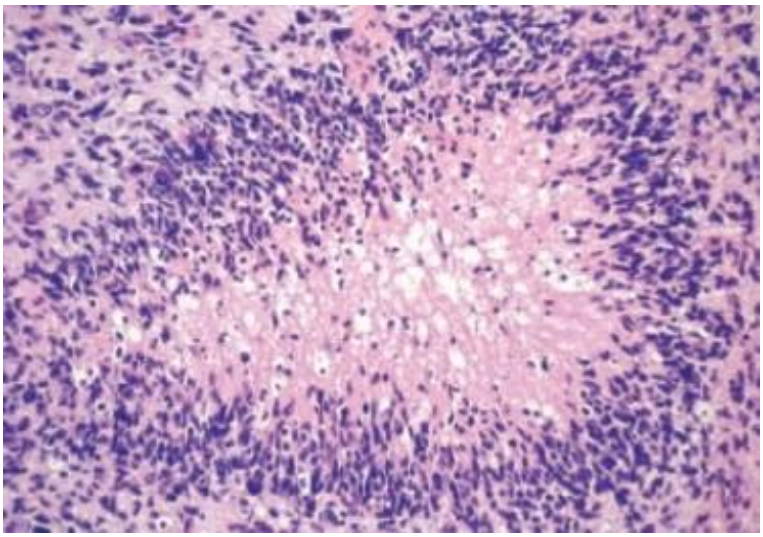


Figure 3 - Histopathologic feature of a medium-sized (200-400µm) pseudopalisade in GBM
Pseudopalisades are the key-features to histologically distinguish GBM from other adult gliomas, presenting central necrosis, central vacuolization, individual dying cells and typical peripheral zone of fibrillarity right inside the feature. (figure adapted from (37))

1.1.6 Genetic and molecular pathology: the key to diagnostics

There are six main intracellular events featuring glioma and the combination of multiple of those oncogenic occurrences are the cause and sustenance of GBM (38): (i) loss of cell cycle control, (ii) genetic instability (iii) overexpression of growth factors and their receptors, (iv) invasion and migration, (v) abnormality of apoptosis and (vi) angiogenesis (38). The genomic profiling which has been used to identify the genetic abnormalities underlying GBM tumorigenesis (39), brought a set of genes whose alterations serve as tumour classification/diagnostic, prognostic and therapy response biomarkers, demonstrating its significant genetic heterogeneity (40,41).

Primary GBMs are mostly characterized by *EGFR* gene-amplification and overexpression, *PTEN* mutations, *TERT* promotor mutations, loss of heterozygosity (LOH) of chromosome 10q and they typically lack *IDH1/2* mutations, which carry the worst prognosis. In contrast, the hallmarks of secondary GBMs include mutations in *IDH1/2*, *TP53* and *ATRX* genes and LOH of 19q, which are biomarkers associated with longer survival (42,41,10,34,19). Paediatric GBMs, a unique malignancy, most commonly feature alterations in *H3F3A*, *ATRX* and *TP53* genes and rarely feature *IDH* mutations (43,41,42).

Table 1 summarizes the most frequently mutated genes in gliomas and represents the list of target regions examined in the present study.

Table 1 - Central features of each gene examined in the present study

GENE	Gene abbreviation	Genomic location; main subcellular locations	Physiological function	Mutation consequence	Glioma type	References
Activin A Receptor Type 1	*ACVR1*	Chromosome 2; plasma membrane, extracellular	Member of the bone morphogenic protein (BMP) signalling pathway: type I BPM receptor, which is activated in response to extracellular BMP, phosphorylating and activating downstream growth	Gain of function: constitutive activation of the BMP signalling pathway, leading to uncontrolled cell proliferation	Pediatric HGGs, mostly related to DIPG (20–32%)	(44,43, 45,46)
Alpha Thalassemia/Mental Retardation Syndrome X-linked	*ATRX*	Chromosome X; nucleus	DNA helicase crucial in chromatin remodelling and maintenance of telomeres	Loss of function: alternative lengthening of telomeres (ALT) phenotype, leading to cell immortalization	Astrocytomas Grade II-III (71%) Oligoastrocytoma (68%) Secondary GBM (57%) *Uncommon* Pediatric GBM (20%); Oligodendrogliomas (14%) *Rare* Primary GBM (4%)	(47,17 48,19)
B-Raf Proto-Oncogene, Serine/Threonine Kinase	*BRAF*	Chromosome 7; nucleus, cytoplasm, plasma membrane	Strongly activates the extracellular signal-regulated/mitogen-activated protein kinase 1 and 2 (MAPK/ERKs) signalling pathway, playing a key-role in cellular division, survival and metabolism	Gain of function: constitutive activation of the MAPK/ERKs signalling pathway, leading to uncontrolled cell proliferation Most frequent BRAF mutations occur at amino acid position 600 (BRAF-V600E)	LGG: Pleomorphic xanthoastrocytoma (50–60%) Ganglioglioma (20%-60%) Pilocytic astrocytoma (10%)	(47,17 49,13)

					Rare in adult HGG (2–5%) but more frequent in paediatric GBMs with closer association to giant-cell GBMs	
Capicua Transcriptional Repressor	*CIC*	Chromosome 19; nucleus	Transcriptional repressor of receptor tyrosine kinase (RTK) signalling pathway	Loss of function: activation of RTK signalling pathway leading to upregulation of genes involved in cell proliferation Associated with 1p/19q codeletion	Oligodendrogliomas Grade II-III (52%) *Rare* Primary GBM (1%)	(50,48)
Epidermal Growth Factor Receptor	*EGFR*	Chromosome 7; plasma membrane, nucleus, endoplasmic reticulum, golgi apparatus, endosome, cytosol, extracellular	Transmembrane receptor which binds to epidermal growth factor (EGF) initiating the EGFR signalling pathway which leads to transcription and cell proliferation	Gain of function: uncontrolled cell proliferation	Primary GBM (40%) Often related to giant-cell GBM	(38,19)
Erb-B2 Receptor Tyrosine Kinase 2	*ERBB2*	Chromosome 17; plasma membrane, nucleus, endosome, cytosol	Member of EGFR pathway whose activation leads to transcription and cell proliferation	Gain of function: constitutive activation of the EGFR signalling pathway, leading to uncontrolled cell proliferation	*	(51,52)
Fibroblast Growth Factor Receptor 1	*FGFR1*	Chromosome 8; plasma membrane, nucleus, cytosol, extracelular	RTK whose extracellular portion interacts with fibroblast growth factors, regulating mitogenesis, proliferation, differentiation, cellular migration and angiogenesis	Gain of function: uncontrolled cell proliferation	GBM Pilocytic astrocytoma Ependymoma	(53,54)

Far Upstream Element Binding Protein 1	*FUBP1*	Chromosome 1; nucleus	Regulates c-Myc proto-oncogene expression by binding to its far-upstream element (FUSE)	Gain of function: stimulation of c-Myc expression leading to uncontrolled cell proliferation	Oligodendrogliomas Grade II-III (30%) *Rare* Primary GBM (2%)	(48,55)
H3 Histone Family Member 3A	*H3F3A* (H3.3)	Chromosome 1; extracellular, nucleus	Nuclear proteins responsible for the complex post-translational epigenetic expression by mediating changes in the DNA heterochromatin structure and conducting the interactions of transcriptional activators and repressors	Loss of function: continued transcription leading to uncontrolled cell proliferation Most frequent mutations are SNVs leading to substitutions G34R/V or K27M	Paediatric and young adult HGG (80% DIPG and 22% other GBMs)	(16,56)
H3 Histone Cluster 1 Family Member B	*HIST1HB* (H3.1)	Chromosome 6; extracellular, nucleus				
Isocitrate Dehydrogenase 1 (NADP+), cytoplasmic	*IDH1*	Chromosome 2; extracellular, peroxisome, cytosol, mitochondrion	IDH family catalyse the oxidative decarboxylation of isocitrate to α-ketoglutarate (α-KG), vital to metabolic processes, while converting NADP+ to NADPH, which protects the cell from reactive oxygen species (ROS)	Preferential affinity for α-KG than for isocitrate leading to a neomorphic activity that mediates the abnormal formation of the oncometabolite hydroxyglutarate (2-HG) as a result of NADPH-dependent reduction of α-KG: - inhibition of multiple α-KG-dependent enzymes leading to epigenetic changes: hypermethylation at large number of genes resulting in gene expression alterations and inactivation of tumour supressors; increased histone methylation at oncogenic sites (H3k27), resulting in a blockage to cell differentiation and uncontrolled cell proliferation - hypersensibility to ROS and oxidative damage Featured by single amino acid substitutions replacing arginine in codons 132 of *IDH1* or 172 of *IDH2*; *IDH1* R132H is the most prevailing mutation in glioma; *IDH1* and *IDH2* mutations are mutually exclusive	Secondary GBM (60–90%) Oligodendroglioma (64–93%) Diffuse astrocytoma (54–100%) Anaplastic astrocytoma (66%) *Rare* Primary GBM (3–7%) Pediatric gliomas	(13,7, 47,17 19,57 21,34)
Isocitrate Dehydrogenase (NADP+) 2, mitochondrial	*IDH2*	Chromosome 15; extracellular, mitochondrion, cytosol				

Neurofibromin 1	*NF1*	Chromosome 17; nucleus, cytosol, plasma membrane	Potential tumour suppressor which is a negative regulator of Ras and mechanistic target of rapamycin (mTOR) signalling pathways	Loss of function: enhancement of Ras/MAPK/ERKs signalling pathway, leading to malignant transformation	GBM	(21,7)
Phosphoinositide -3-Kinase Catalytic Subunit Alpha	*PIK3CA*	Chromosome 3; cytosol, plasma membrane	Member of the oncogenic receptor tyrosine kinase/phosphoinositide 3'-kinase/a serine-threonine kinase (RTKs/PI3K/Akt) signalling pathway which regulates cell growth, survival, proliferation, motility and adhesion	Gain of function: constitutive activation of the RTKs/PI3K/Akt signalling pathway, leading to uncontrolled cell proliferation, cell invasion and blood vessel formation	GBM	(7,58, 59)
Phosphoinositide -3-Kinase Regulatory Subunit 1	*PIK3R1*	Chromosome 5; plasma membrane, nucleus, cytosol				
Protein Phosphatase, Mg^{2+}/Mn^{2+} Dependent 1D	*PPM1D*	Chromosome 17; nucleus	Member of the PP2C family of serine/threonine protein phosphatases which negatively regulates cell stress response pathways as a negative regulator of TP53 tumour suppressor	Gain of function: inhibition of TP53 functions	*	(60,61)
Phosphatase And Tensin Homolog	*PTEN*	Chromosome 10; plasma membrane, nucleus, cytosol, mitochondrion, extracellular.	Tumour suppressor which negatively regulates the oncogenic RTKs/PI3K/Akt signalling pathway by antagonizing PI3K function	Loss of function: activation of the RTKs/PI3K/Akt signalling pathway, leading to uncontrolled cell proliferation, cell invasion and blood vessel formation *PTEN* deletion associated with LOH of chromosome 10q	Primary GBM (24-34%) Anaplastic astrocytoma (18%) *Rare* Secondary GBM (4%)	(38,62 21,7)
Telomerase Reverse Transcriptase	*TERT*	Chromosome 5; mitochondrion, nucleus	Ribonucleoprotein polymerase that regulates the length of telomere ends of chromosomes Inactive or very low activity in most of adult cells	Gain of function: cellular immortalization and proliferation Most frequent *TERT* mutations occur at positions 228 and 250 in the TERT promotor	Primary GBM (70-80%) Oligodendroglioma (70-80%)	(19,13, 17,47, 14)

					Uncommon Secondary GBM (5-26%) *Rare* Pediatric GBM	
Tumour Protein P53	*TP53*	Chromosome 17; mitochondrion, nucleus, endoplasmic reticulum, cytosol	Tumour suppressor and transcription factor which coordinates cell responses to cellular stresses by regulating genes involved in cell division (downregulation) or apoptosis (upregulation)	Loss of function: apoptosis inhibition and uncontrolled cell proliferation	Secondary GBM (60%-70%) Primary GBM (25%-30%) More frequent in younger patients	(7,19, 62,14)

*Not yet specified (genomic locations and main subcellular locations were collected resorting http://www.genecards.org/)
DIPG – diffuse intrinsic pontine glioma
HGG – High-grade glioma
LGG – Low-grade glioma

1.2 Prioritizing molecular markers in the diagnostics of brain tumours

For the past century and until recently, the standard principles used by neuropathologists towards the diagnostics and classification of brain tumours have been based predominantly on histological concepts which categorized tumours according to their microscopic similarities with the origin cells and supposed differentiation levels, defined in the 2007 version of the WHO classification (16,14). Despite this histologic-based classification system resulted of constant updates and evolved over the years, it carries several limitations, such as inconclusive histological features due to microscopic similarities between different entities, inter-observer variability and insufficient or non-representative tissue sampling. Hence, it is known that patients with morphologically identical neoplasms may experience different clinical outcomes and treatment responses due to distinct underlying genetic features of the tumours. As such, there is increased evidence that histologically ambiguous gliomas may be more precisely classified resorting genetic markers (16,18,63). A notorious example is the controversial group of oligoastrocytomas which can turn out to fit into either diffuse astrocytomas or oligodendrogliomas, depending on three genetic markers: *IDH1/2*, 1p/19q codeletion and loss of *ATRX* (18), as displayed in Figure 5.

Emphasizing the increasing impact of molecular information, the WHO 2016 new classification of the CNS tumours incorporated molecular parameters into the tumour classification criteria, breaking with the previous approach based entirely on microscopy (8,14). Presently, a CNS tumour classification consists of a histopathological name followed by the genetic features (e.g. Glioblastoma, *IDH*-mutant), as seen in Figure 4. When tumours have not been fully tested for the relevant genetic parameters, a diagnostic designation of NOS (not otherwise specified) may be acceptable (14).

Hence, GBMs are subdivided in the 2016 WHO classification of CNS tumours into (a) glioblastoma, *IDH*-wildtype which corresponds to the clinically defined primary GBM (90% of cases); (b) glioblastoma, *IDH*-mutant, which corresponds to the secondary GBM; (c) glioblastoma, NOS, when full *IDH* assessment cannot be performed (Figures 4, 5 and 6) (14).

Each glioma subtype has its feature molecular profile, differing from each other and allowing to perform a more accurate diagnostics and to explore new potential targeted therapies (14).

WHO classification of tumours of the central nervous system

Diffuse astrocytic and oligodendroglial tumours	
Diffuse astrocytoma, IDH-mutant	9400/3
Gemistocytic astrocytoma, IDH-mutant	9411/3
Diffuse astrocytoma, IDH-wildtype	*9400/3*
Diffuse astrocytoma, NOS	9400/3
Anaplastic astrocytoma, IDH-mutant	9401/3
Anaplastic astrocytoma, IDH-wildtype	*9401/3*
Anaplastic astrocytoma, NOS	9401/3
Glioblastoma, IDH-wildtype	9440/3
Giant cell glioblastoma	9441/3
Gliosarcoma	9442/3
Epithelioid glioblastoma	*9440/3*
Glioblastoma, IDH-mutant	9445/3*
Glioblastoma, NOS	9440/3
Diffuse midline glioma, H3 K27M-mutant	9385/3*
Oligodendroglioma, IDH-mutant and 1p/19q-codeleted	9450/3
Oligodendroglioma, NOS	9450/3
Anaplastic oligodendroglioma, IDH-mutant and 1p/19q-codeleted	9451/3
Anaplastic oligodendroglioma, NOS	*9451/3*
Oligoastrocytoma, NOS	*9382/3*
Anaplastic oligoastrocytoma, NOS	*9382/3*
Other astrocytic tumours	
Pilocytic astrocytoma	9421/1
Pilomyxoid astrocytoma	9425/3
Subependymal giant cell astrocytoma	9384/1
Pleomorphic xanthoastrocytoma	9424/3
Anaplastic pleomorphic xanthoastrocytoma	9424/3
Ependymal tumours	
Subependymoma	9383/1
Myxopapillary ependymoma	9394/1
Ependymoma	9391/3
Papillary ependymoma	9393/3
Clear cell ependymoma	9391/3
Tanycytic ependymoma	9391/3
Ependymoma, *RELA* fusion-positive	9396/3*
Anaplastic ependymoma	9392/3
Other gliomas	
Chordoid glioma of the third ventricle	9444/1
Angiocentric glioma	9431/1
Astroblastoma	9430/3
Choroid plexus tumours	
Choroid plexus papilloma	9390/0
Atypical choroid plexus papilloma	9390/1
Choroid plexus carcinoma	9390/3
Neuronal and mixed neuronal-glial tumours	
Dysembryoplastic neuroepithelial tumour	9413/0
Gangliocytoma	9492/0
Ganglioglioma	9505/1
Anaplastic ganglioglioma	9505/3
Dysplastic cerebellar gangliocytoma (Lhermitte-Duclos disease)	9493/0
Desmoplastic infantile astrocytoma and ganglioglioma	9412/1
Papillary glioneuronal tumour	9509/1
Rosette-forming glioneuronal tumour	9509/1
Diffuse leptomeningeal glioneuronal tumour	
Central neurocytoma	9506/1
Extraventricular neurocytoma	9506/1
Cerebellar liponeurocytoma	9506/1
Paraganglioma	8693/1
Tumours of the pineal region	
Pineocytoma	9361/1
Pineal parenchymal tumour of intermediate differentiation	9362/3
Pineoblastoma	9362/3
Papillary tumour of the pineal region	9395/3
Embryonal tumours	
Medulloblastomas, genetically defined	
Medulloblastoma, WNT-activated	9475/3*
Medulloblastoma, SHH-activated and *TP53*-mutant	9476/3*
Medulloblastoma, SHH-activated and *TP53*-wildtype	9471/3
Medulloblastoma, non-WNT/non-SHH	9477/3*
Medulloblastoma, group 3	
Medulloblastoma, group 4	
Medulloblastomas, histologically defined	
Medulloblastoma, classic	9470/3
Medulloblastoma, desmoplastic/nodular	9471/3
Medulloblastoma with extensive nodularity	9471/3
Medulloblastoma, large cell / anaplastic	9474/3
Medulloblastoma, NOS	9470/3
Embryonal tumour with multilayered rosettes, C19MC-altered	9478/3*
Embryonal tumour with multilayered rosettes, NOS	9478/3
Medulloepithelioma	9501/3
CNS neuroblastoma	9500/3
CNS ganglioneuroblastoma	9490/3
CNS embryonal tumour, NOS	9473/3
Atypical teratoid/rhabdoid tumour	9508/3
CNS embryonal tumour with rhabdoid features	9508/3
Tumours of the cranial and paraspinal nerves	
Schwannoma	9560/0
Cellular schwannoma	9560/0
Plexiform schwannoma	9560/0

Figure 4 - The 2016 WHO classification of the CNS tumours
(figure origin:(14))

Figure 4 - continued

Melanotic schwannoma	9560/1
Neurofibroma	9540/0
Atypical neurofibroma	9540/0
Plexiform neurofibroma	9550/0
Perineurioma	9571/0
Hybrid nerve sheath tumours	
Malignant peripheral nerve sheath tumour	9540/3
Epithelioid MPNST	9540/3
MPNST with perineurial differentiation	9540/3
Meningiomas	
Meningioma	9530/0
Meningothelial meningioma	9531/0
Fibrous meningioma	9532/0
Transitional meningioma	9537/0
Psammomatous meningioma	9533/0
Angiomatous meningioma	9534/0
Microcystic meningioma	9530/0
Secretory meningioma	9530/0
Lymphoplasmacyte-rich meningioma	9530/0
Metaplastic meningioma	9530/0
Chordoid meningioma	9538/1
Clear cell meningioma	9538/1
Atypical meningioma	9539/1
Papillary meningioma	9538/3
Rhabdoid meningioma	9538/3
Anaplastic (malignant) meningioma	9530/3
Mesenchymal, non-meningothelial tumours	
Solitary fibrous tumour / haemangiopericytoma**	
Grade 1	8815/0
Grade 2	8815/1
Grade 3	8815/3
Haemangioblastoma	9161/1
Haemangioma	9120/0
Epithelioid haemangioendothelioma	9133/3
Angiosarcoma	9120/3
Kaposi sarcoma	9140/3
Ewing sarcoma / PNET	9364/3
Lipoma	8850/0
Angiolipoma	8861/0
Hibernoma	8880/0
Liposarcoma	8850/3
Desmoid-type fibromatosis	8821/1
Myofibroblastoma	8825/0
Inflammatory myofibroblastic tumour	8825/1
Benign fibrous histiocytoma	8830/0
Fibrosarcoma	8810/3
Undifferentiated pleomorphic sarcoma / malignant fibrous histiocytoma	8802/3
Leiomyoma	8890/0
Leiomyosarcoma	8890/3
Rhabdomyoma	8900/0
Rhabdomyosarcoma	8900/3
Chondroma	9220/0
Chondrosarcoma	9220/3
Osteoma	9180/0
Osteochondroma	9210/0
Osteosarcoma	9180/3
Melanocytic tumours	
Meningeal melanocytosis	8728/0
Meningeal melanocytoma	8728/1
Meningeal melanoma	8720/3
Meningeal melanomatosis	8728/3
Lymphomas	
Diffuse large B-cell lymphoma of the CNS	9680/3
Immunodeficiency-associated CNS lymphomas	
AIDS-related diffuse large B-cell lymphoma	
EBV-positive diffuse large B-cell lymphoma, NOS	
Lymphomatoid granulomatosis	9766/1
Intravascular large B-cell lymphoma	9712/3
Low-grade B-cell lymphomas of the CNS	
T-cell and NK/T-cell lymphomas of the CNS	
Anaplastic large cell lymphoma, ALK-positive	9714/3
Anaplastic large cell lymphoma, ALK-negative	9702/3
MALT lymphoma of the dura	9699/3
Histiocytic tumours	
Langerhans cell histiocytosis	9751/3
Erdheim–Chester disease	9750/1
Rosai–Dorfman disease	
Juvenile xanthogranuloma	
Histiocytic sarcoma	9755/3
Germ cell tumours	
Germinoma	9064/3
Embryonal carcinoma	9070/3
Yolk sac tumour	9071/3
Choriocarcinoma	9100/3
Teratoma	9080/1
Mature teratoma	9080/0
Immature teratoma	9080/3
Teratoma with malignant transformation	9084/3
Mixed germ cell tumour	9085/3
Tumours of the sellar region	
Craniopharyngioma	9350/1
Adamantinomatous craniopharyngioma	9351/1
Papillary craniopharyngioma	9352/1
Granular cell tumour of the sellar region	9582/0
Pituicytoma	9432/1
Spindle cell oncocytoma	8290/0
Metastatic tumours	

The morphology codes are from the International Classification of Diseases for Oncology (ICD-O) (742A). Behaviour is coded /0 for benign tumours; /1 for unspecified, borderline, or uncertain behaviour; /2 for carcinoma in situ and grade III intraepithelial neoplasia; and /3 for malignant tumours. The classification is modified from the previous WHO classification, taking into account changes in our understanding of these lesions.
*These new codes were approved by the IARC/WHO Committee for ICD-O. *Italics*: Provisional tumour entities. **Grading according to the 2013 *WHO Classification of Tumours of Soft Tissue and Bone*

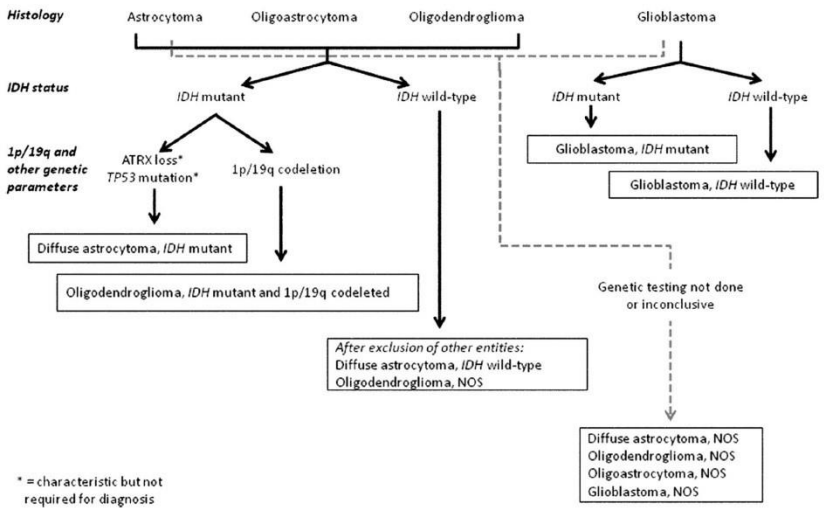


Figure 5 - Simplified algorithm for the diffuse gliomas integrated classification
(figure origin:(14))

"The use of "integrated" phenotypic and genotypic parameters for CNS tumour classification adds a level of objectivity that has been missing from some aspects of the diagnostic process in the past." (14)

	IDH-wildtype glioblastoma	IDH-mutant glioblastoma
Synonym	Primary glioblastoma, IDH-wildtype	Secondary glioblastoma, IDH-mutant
Precursor lesion	Not identifiable; develops de novo	Diffuse astrocytoma Anaplastic astrocytoma
Proportion of glioblastomas	~90%	~10%
Median age at diagnosis	~62 years	~44 years
Male-to-female ratio	1.42:1	1.05:1
Mean length of clinical history	4 months	15 months
Median overall survival		
Surgery + radiotherapy	9.9 months	24 months
Surgery + radiotherapy + chemotherapy	15 months	31 months
Location	Supratentorial	Preferentially frontal
Necrosis	Extensive	Limited
TERT promoter mutations	72%	26%
TP53 mutations	27%	81%
ATRX mutations	Exceptional	71%
EGFR amplification	35%	Exceptional
PTEN mutations	24%	Exceptional

Figure 6 - Key features of *IDH*-wildtype and *IDH*-mutant GBMs
(figure adapted from (14))

Thus, this new integrated brain tumour classification is expected to lead to greater diagnostic certainty, enhanced patient management and more accurate determinations of prognosis and treatment response, highlighting the importance of the molecular parameters in these matters.

Additionally, despite several genetic alterations can be already detected resorting to immunohistochemistry or FISH, including *IDH1* R132H, *ATRX* loss, *H3* K27M and 1p19q codeletion (11), DNA sequencing is imperative to allow a richer analysis capable of fully respond to the increasing requirements for better molecular and clinical understanding.

1.3 Next-Generation Sequencing (NGS) pathway towards molecular diagnostics

Succeeding the increased importance of molecular parameters in the clinical outcomes, an extensive molecular profiling is crucial for tumours of the CNS in order to improve their management. Next-Generation Sequencing (NGS) technology answers to this require, providing the opportunity to analyse a large set of molecular targets in a massively paralleled approach, still granting high accuracy and sensitivity because of its high sequencing coverage (64,63).

1.3.1 The evolution of genomic technologies: DNA sequencing timeline

In 1944, Avery McLeod and McCarthy showed that the DNA was the material from which genes were composed and since that legendary experiment the knowledge regarding DNA is constantly growing (65). In 1953, Watson and Crick made the next significant discovery in the genetics field proposing a double helix as the structure model of DNA, which is still the central principle of molecular biology (65,66). Sequencing technology emerged as the set of methods to determine the primary structure of DNA, allowing to learn about the order of the four nucleotides (A, C, G, T) in the DNA, which remains the growing interest in genetics (65). Bacteriophage Phi-X174 was the first organism whose genome was completely sequenced, in 1977, resulting from the development of the first-generation DNA sequencing technique: Sanger's 'chain-terminator' or di-deoxy method. This was the sequencing gold standard until the late 2000's and it is still one of the most reliable methods to establish molecular diagnostics, along with pyrosequencing technology, later developed in 1996 (66,65,67).

As a result of enthusiastic massive sequencing projects (e.g. human genome), it was fundamental to develop a new technology which carried less costs and required less time to collect the sequences and hereby emerged the Next-Generation Sequencing (NGS) (65,67): "*a high-performance technology based on the parallelization of the sequencing process, resulting in the reading of thousands or millions of sequences simultaneously"* (65). It was in 2004 when the first NGS sequencer, 454 Roche (based on pyrosequencing technology), was commercialized and ever since the increased research in NGS has launched eight large massive sequencing platforms to the market and the NGS associated costs have been continuously decreasing (65,67).

1.3.2 NGS: a technology with unique features

Next-Generation Sequencing (NGS), including second- and third-generation sequencing technologies, universally refers to those technologies which are able to massively parallel sequence millions of DNA templates, allowing to analyse a large number of samples simultaneously. This demonstrates the evolution of the conventional sequencing techniques in which only one or a few short fragments of DNA previously amplified by Polymerase Chain Reaction (PCR) could be sequenced per tube. This great progress was allowed not only by innovation in sequencing chemistries but also by better imaging, microfabrication and information technology (68,67,69).

1.3.2.1 NGS landscape

Currently, this advanced sequencing technology is subdivided in two major categories: short- and long-read sequencing (68). According to the DNA sequencing method, there are several NGS sequencers available (Table 2). The eight distinct platforms presently available diverge from each other concerning not only the general parameters such as preparation of the templates for sequencing, the sequencing chemistry itself and the detection system utilized (70) but also specific intrinsic parameters (equipment, accuracy, numbers of readings and, consequently, cost for each sequencing reaction) (71).

Table 2 - NGS sequencers presently available in the market

Short-read sequencing technology can be performed through either sequencing by ligation or sequencing by synthesis. Sequencing by synthesis sequencers are those with generally better relationship between cost ($10-100/Gb, excepting for 454 Roche which is the most expensive of the sequencers with a cost of $10.000/Gb), accuracy (99-99.9%) and time of run (2h-3days). Illumina is currently the most frequently used platform because of its lowest cost associated with highest number of bases read per run. Sequencing by ligation sequencers, despite being as cheap and as accurate as Illumina's ($10/Gb; 99.9%), have the longer run time (7 days) and the lowest read length (75 bp) comparing to all other sequencers. Long-read sequencing sequencers are associated with higher costs ($600-1000/Gb) and lower accuracy (90%) due to their highest read lengths in line of the thousands (65,67,70,69). (adapted from: (65))

Methods of DNA Sequencing			***Platforms***
Short-read sequencing	Sequencing by ligation		AB SOLiD (Thermo Fisher) Complete Genomics (BGI)
	Sequencing by synthesis	Cyclic reversible termination (CRT)	Illumina GeneReader (Qiagen)
		Single-nucleotide addition (SNA)	454 Roche Ion Torrent
Long-read sequencing	Single-molecule real-time long read sequencing		Pacific Bioscience Oxford Nanopore

1.3.2.2 Illumina's platform

The short-read sequencing method is currently the most frequently used approach due to its low costs per Gb and high accuracy (67). As such, the sequencers most extensively used in the development of massive sequencing projects, including diagnostic purposes, are Illumina's (MiSeq/HiSeq/NextSeq), not only due to the previously referred features associated with short-reading sequencing method but also because of the great diversity of equipment present in the market adapted to the needs of each project (65,41). As general features, Illumina's platform have a mean read length of 300 bp, an accuracy of 99,9%, a run time of 3 days and it processes 1800 Gb per run with a symbolic cost of $10/Gb (65). Basically, Illumina's platform performs DNA sequencing resorting to fluorescence-labelled nucleotide analogues which act as reversible terminators of the amplification reaction, differing from the Sanger sequencing whose DNA polymerization stoppage is irreversible. This NGS technology has a unique feature: it carries out the clonal amplification *in situ* by bridge PCR, joining the DNA templates to immobilized primers present on a solid surface (flow cell) then generating clusters of DNA composed of equal molecules and the fluorescence emission wavelength and intensity directly recorded during sequencing process are used to identify the base incorporated (65,72).

1.3.2.2 Targeted-sequencing

Different approaches can be pursued according to the goal of the project in mind – whole genome, whole exome or targeted sequencing. As the first both are expensive, time-consuming and difficult to perform with the small amounts of DNA (e.g. brain biopsy), targeted sequencing (or re-targeted, when the genes designed to assess were previously studied by the conventional methods) is the method of choice for most clinical applications (e.g. diagnostics, prognostics and predict treatment response), allowing to study the genetic alterations of a neoplastic tissue (63,67). Targeted sequencing is performed resorting to gene panels which can be purchased with preselected content or custom designed including the genomic regions of interest (ROIs). As such, they have the ability of detecting genetic variations such as single-nucleotide variations (SNVs), insertions and deletions, copy number changes and gene fusions, in a cost-efficient and a high-coverage of ROIs manner (63,67). There are various targeted NGS panels available in the market, but not even one is designed to specifically target the critical genetic alterations in the tumours of the CNS (63).

1.3.3 Prevailing pros and cons of NGS technology for glioma diagnostic routine

NGS technology brought three significant enhancements from the standard traditional Sanger sequencing and pyrosequencing methods: (i) NGS do not require previous amplification of DNA fragments by PCR or bacterial cloning; (ii) NGS can in one sequencing reaction analyse multiple genes/ROIs of multiple samples, handing out the need of hundreds of sequencing reactions which would be needed if one would use conventional technologies (analyse one gene from one sample per sequencing reaction); (iii) NGS outputs are directly detected oppositely to Sanger sequencing which requires further electrophoresis (69). As such, short-read sequencing platforms, e.g. Illumina's, provide a cost- and time efficient alternative to currently used conventional sequencing techniques, which follow the single-gene-approach, and, resorting to the gene panels approach, it could theoretically be implemented for routine cancer diagnostics (67,41).

However, NGS has not yet been installed in routine brain tumour diagnostics and so neither has replaced the established traditional methods due to the lack of sufficiently powered studies providing evidence that NGS is reliable enough for this purpose (64,41). The paucity of evidence is the consequence of the remaining great challenges of this new technology: (a) how to most efficiently filter all data outputs in order to select the important mutations and distinguish them from potentially false-positives, e.g. result of poor-quality DNA artefacts and (b) how to manage all information and non-traditional sorts of findings, such as novel mutations which can or cannot be clinically significant (64,67,73).

Hence, and being the major guide of the present study, it is imperative to deeply study all NGS method-related features in order to determine its leverage when performing the molecular diagnostics of brain tumours comparing to the gold-standard classical technologies currently used.

2 SCOPE OF THE STUDY

This study is focused on the molecular diagnostics of brain tumours and it was developed at the Brain Tumour Reference Centre of Germany.

For the first time, WHO 2016 Classification of Brain Tumours includes molecular parameters in addition to histology to establish the diagnostics of many tumour entities (14). The more important molecular diagnostics of brain tumours get, the more crucial is to improve the pathways to achieve it. Single-gene approach, such as pyrosequencing or Sanger sequencing technologies, is the most used classical way to determine the molecular diagnostics of brain tumours.

In this study, the main goal is to apply and verify if NGS technology could effectively replace the classical methods to establish glioma's molecular diagnostics. In this purpose, two approaches were scrutinized:

1) *Detection of "old mutations"* – assessment of NGS ability to correctly detect well-known and described mutations, such as hotspot mutations in codon 132 of *IDH1*, in codon 600 of *BRAF* and in codons 27 and 34 of *H3F3A*, through further validation with pyrosequencing single-gene analysis. Also, six glioma DNA samples with several different molecular diagnostics already performed by pyrosequencing analysis were added in the study to evaluate the accuracy of NGS technology concerning the detection of the mutations already identified with the classical method.
2) *Analysis of "new mutations"* – NGS mutation results are endless. Many "new" (not known or described) mutations are displayed in NGS outputs, but are they real or just artefacts? In this concern, an extensive data analysis of NGS outputs through a good filtering process considering all technical and biological factors shall be done to allow a selection of the most promising novel mutations to further evaluate and possibly validate with the gold-standard Sanger sequencing technology.

Thus, an overall analysis of NGS technology characteristics was performed to appraise if NGS pros mute NGS cons regarding those of the classical technologies currently performed to establish the molecular diagnostics of brain tumours.

3 MATERIALS AND METHODS

3.1 Materials

3.1.1 FFPE DNA Samples

Twenty-four DNA samples from GcGBM were selected to analyse: twelve from adult patients and the other twelve from children patients; in addition, six samples from other gliomas with molecular diagnostics already performed and known were used as controls (n = 30) (Table 3).

All of the DNA samples were formalin-fixed paraffin-embedded (FFPE) preserved from surgically removed CNS tumours and the DNA was extracted using the QIAamp DNA FFPE Tissue Kit.

Table 3 - Samples' ID correspondences
In this table it is represented each DNA sample ID (**bold**) and the correspondent number for each (sample X) further used in this study to simplify the terminology and the graphing. The first 24 samples (1-24) are from GcGBM and the last six (25-30) are from other glioma samples which were used as controls.

Sample ID			
585	Sample 1	**4121**	Sample 16
588	Sample 2	**4122**	Sample 17
748	Sample 3	**4125**	Sample 18
2489	Sample 4	**4129**	Sample 19
4099	Sample 5	**4142**	Sample 20
4100	Sample 6	**4153**	Sample 21
4101	Sample 7	**4155**	Sample 22
4102	Sample 8	**4158**	Sample 23
4107	Sample 9	**4160**	Sample 24
4109	Sample 10	**R-78610**	Sample 25
4110	Sample 11	**P-3879**	Sample 26
4111	Sample 12	**R-78222**	Sample 27
4118	Sample 13	**R-78653**	Sample 28
4119	Sample 14	**R-77342**	Sample 29
4120	Sample 15	**R-77376**	Sample 30

3.1.2 Primers

Table 4 - Primers for amplification of glioma gDNA for hot spot mutations analysis

Gene (locus)	*Primer sequence*
BRAF (codon 600)	fw: 2708 5’ GAAGACCTCACAGTAAAAATAG 3’ rev: 2709 5’ Biotin-ATAGCCTCAATTCTTACCATCC3’
IDH1 (codon 132)	fw: 2134 5’ CACCATACGAAATATTCTGG 3’ rev: 2137 5’ Biotin-CAACATGACTTACTTGATCC 3'
H3F3A (codon 27 and 34)	fw: 2711 5’ TGTTTGGTAGTTGCATATGG 3’ rev: 2712 5’ Biotin-TACAAGAGAGACTTTGTCC 3’

Table 5 - Primers for pyrosequencing of amplified glioma gDNA for hot spot mutations analysis

Gene (locus)	*Primer sequence*
BRAF (codon 600)	ps: 2710 5’ AGGTGATTTTGGTCTAGCTACAG 3’
IDH1 (condon 132)	ps: 2136 5’ GTGAGTGGATGGGTAAAACC 3’
H3F3A (codon 27 and 34)	ps: 2705 5’ CAAAAGCCGCTCGCA 3’

3.1.3 Molecular biological kits

3.1.3.1 Next-Generation Sequencing (NGS)

a) Input DNA quantification: QUBIT

Qubit® dsDNA BR Assay Kit Invitrogen (Darmstadt, Germany)

Qubit® dsDNA BR Reagent

Qubit® dsDNA BR Buffer

Qubit® dsDNA BR Standard #1

Qubit® dsDNA BR Standard #2

b) Libraries preparation

Illumina® TruSeq® Custom Amplicon Low Input Library Prep Kit (96 samples) Illumina Inc (Ense-Höingen, Germany)

Table 6 - Low Input Library Prep Kit contents

Box 1/3: pre-PCR	Box 2/3: pre-PCR	Box 3/3: post-PCR
ACP3	SS1	HT1
ELE	SPB	LNA1
ELB	OHS2	LNW1
EDP	SW1	LNS2
EMM	RS1	RSB
2800M	LNB1	

Illumina® TruSeq® Custom Amplicon Index Kit (384 samples) Illumina Inc (Ense-Höingen, Germany)

Table 7 - Index Kit contents

i5 Index Adapter	*Sequence*	*Index replacement caps*
A501	TGAACCTT	A5XX index tube caps – white
A502	TGCTAAGT	
A503	TGTTCTCT	
A504	TAAGACAC	
A505	CTAATCGA	
A506	CTAGAACA	
A507	TAAGTTCC	
A508	TAGACCTA	
i7 Index adapter	*Sequence*	*Index replacement caps*
A701	ATCACGAC	A7XX index tube caps - orange
A702	ACAGTGGT	
A703	CAGATCCA	
A704	ACAAACGG	
A705	ACCCAGCA	
A706	AACCCCTC	
A707	CCCAACCT	
A708	CACCACAC	
A709	GAAACCCA	
A710	TGTGACCA	
A711	AGGGTCAA	
A712	AGGAGTGG	

c) Assessing libraries' preparation success

Agilent DNA 1000 Kit — Agilent Technologies (Waldbronn, Germany)

- DNA Chips
- Electrode Cleaner
- Syringe
- Spin Filters
- DNA Dye Concentrate
- DNA Gel Matrix
- DNA Ladder
- DNA Markers

3.1.3.2 Amplification of gDNA (PCR)

PyroMark PCR Kit — Qiagen (Hilden, Germany)

- Mater mix 2x
- CoralLoad concentrate 10x
- H_2O

3.1.3.3 Pyrosequencing

PyroMark GoldReagents Kit — Qiagen (Hilden, Germany)

- PyroMark Annealing Buffer
- PyroMark Binding Buffer, pH 7,6
- PyroMark Denaturation Solution
- PyroMark Enzyme Mixture
- PyroMark Nucleotides (dATP, dCTP, dGTP, dTTP)
- PyroMark Substract Mixture
- PyroMark Wash Buffer, pH 7,6

3.1.3.4 Sanger Sequencing

Crude PCR products prepaid Kit 96 Clean-ups of crude PCR products 96 sequencing reactions Barcode labelled green 96-well plate Twelve 8-cap strips for sealing Shipping box Pre-paid, pre-addressed, shipping envelope	Eurofins Genomics (Ebersberg, Germany)

3.1.4 Other chemical substances and solutions

Table 8 - Other chemical substances and solutions used in this project

Boric acid	Roth Chemie GmbH (Karlsruhe, Germany)
EDTA (0,5M; pH8,0)	Roth Chemie GmbH (Karlsruhe, Germany)
Ethanol absolute	PanReac AppliChem (Darmstadt, Germany)
HPLC Water	Avantor Performance Materials B.V. (Deventer, Netherlands)
Midori green	NIPPON Genetics EUROPE GmbH (Dueren, Germany)
NaOH 10N	Merck KGaA (Darmstadt, Germany)
peqGOLD Universal Agarose	PEQLAB Biotechnologie GmbH (Erlangen, Germany)
pUC19 DNA ladder	Thermo Fisher Scientific (Schwerte, Germany)
Streptavidin-Sepharose High Performance Beads	GE-Healthcare (Solingen, Germany)
Tris base	Roth Chemie GmbH (Karlsruhe, Germany)

3.1.5 Technical equipment and other materials

Table 9 - Technical equipment and other materials used in this project

Technical equipment	
Agilent 2100 bioanalyzer	Agilent Technologies (Waldbronn, Germany)
Centrifuge HERAEUS RS 3.0	Thermo Fisher Scientific (Schwerte, Germany)
CarlRoth™ Mini-centrifuges	Thermo Fisher Scientific (Schwerte, Germany)
Gel electrophoresis chamber (Sub-Cell® GT Cell)	Biorad (Munich, Germany)
Heat block	Techne, UK
INFINITY^{T} 3000 Imaging Systems	PeqLab Biotechnologie GmbH (Erlangen, Germany)
Monoshake Microplate Shaker	Variomag®-USA (Daytona Beach, Florida, USA)
Quantum ST4 3000	Montreal Biotech (Montreal, Canada)
NGS Sequencer (Illumina® MiSeq System)	Illumina Inc (Ense-Höingen, Germany)
PCR machine (thermal cycler Biometra TRIO)	Biometra (Göttingen, Germany)
pH meter	Mettler Toledo (Giessen, Germany)
PowerPac™ Basic Power Supply	Biorad (Munich, Germany)
PyroMark Q24 Instrument, Biotage	Qiagen (Hilden, Germany)
PyroMark Q24 Vacuum Workstation	Qiagen (Hilden, Germany)
PyroMark Q24 Plate	Qiagen (Hilden, Germany)

PyroMark Q24 Cartridge	Qiagen (Hilden, Germany)
Qubit® 2.0 Fluorometer	Thermo Fisher Scientific (Schwerte, Germany)
Thermoblock	Biometra (Göttingen, Germany)
Vortexer	VWR International (Darmstadt, Germany)
Weigh balance	Kern (Balingen, Germany)
Other materials	
Eppendorf® tubes 0.5, 1.5, 2 mL	Eppendorf (Hamburg, Germany)
Falcon™ Conical Centrifuge Tubes 15,50 mL	Thermo Fisher Scientific (Schwerte, Germany)
Magnetic stand-96	Thermo Fisher Scientific (Schwerte, Germany)
Microseal® 'A' PCR Plate Sealing Film	Bio-Rad Laboratories GmbH (Munich, Germany)
Microseal® 'B' PCR Plate Sealing Film	Bio-Rad Laboratories GmbH (Munich, Germany)
MiSeq Flow Cell	Illumina Inc (Ense-Höingen, Germany)
MiSeq Reagent Cartridge	Illumina Inc (Ense-Höingen, Germany)
PCR 8-tube strips	Thermo Fisher Scientific (Schwerte, Germany)
96-well midi plates	Thermo Fisher Scientific (Schwerte, Germany)
96-well skirted PCR plate	Thermo Fisher Scientific (Schwerte, Germany)

3.1.6 Software and Databases

Table 10 - Software and databases used in this project

Catalogue of Somatic Mutations in Cancer (COSMIC)	Wellcome Trust Sanger Institute (Hinxston, UK)
GeneCards® Human Gene Database	Weizmann Institute of Science (Rehovot, Israel)
Illumina Experiment Manager 1.9	Illumina Inc (Ense-Höingen, Germany)
Illumina Sequencing Analysis Viewer 1.8.37	Illumina Inc (Ense-Höingen, Germany)
Illumina VariantStudio Variant Analysis Software 2.2	Illumina Inc (Ense-Höingen, Germany)
Illustrator for Biological Sequences (IBS) 1.0.2	The Cuckoo Workgroup (China)
Integrative Genomics Viewer (IGV) 2.3	Broad Institute (Cambridge, USA)
PyroMark Q24 Software	Qiagen (Hilden, Germany)
Quantum-Capt Image Software	Montreal Biotech (Montreal, Canada)
UCSC Genome Browser	University of California (Santa Cruz, California, USA)

3.2 Methods

3.2.1 Targeted Next-Generation Sequencing (Illumina® MiSeq System)

3.2.1.1 Design of the glioma-tailored customized gene panel

The gene panel for the assay was previously prepared. It consisted of 518 amplicons covering 19 glioma-related genes: *FUBP1, H3F3A, ACVR1, IDH1, PIK3CA, PIK3R1, TERT, HIST1H3B, BRAF, EGFR, FGFR1, PTEN, IDH2, ERBB2, NF1, TP53, PPM1D, CIC* and *ATRX*, including the most commonly mutated genes in gliomas (42).

3.2.1.2 Library preparation

The major step of this technology is the preparation of a good quality library for further massively parallel sequencing. In this work, libraries were prepared resorting to *Illumina® TruSeq® Custom Amplicon Low Input Library Prep Reference Guide*, according to the manufacturer's instructions. Each step is briefly described below.

I. Quantification and dilution of gDNA samples

DNA samples were quantified using a fluorometric method (QUBIT). 250ng of DNA were diluted in 1µl SS1, to a final volume of 5µl (water was added when needed).

II. Hybridization of oligo pool

Then, the hybridization of the custom oligo pool containing upstream and downstream oligos specific to the target regions of interest (Figure 7) was performed leading to replicates that increase confidence in variant calls.
Briefly, in a 96-well PCR plate (Thermo Fisher Scientific, Germany), 2,5µl CAT, 2,5µl RS1 and 15µl OHS2 were added per sample well. A no template control was performed adding 5µl RS1, 5µl CAT and 15µl OHS2; also, an assay control was executed, adding 2µl 2800M, 2µl RS1, 5µl ACP3 and 15µl OHS2. For 25µl reaction volume, the hybridization program run consisted of:

1. 95°C for 3minutes
2. from 90°C, decrease by 0.5°C, hold for 30seconds, ramp at 0.1°C per second (60x)
3. from 60°C, decrease by 0.5°C, hold for 1minute, ramp at 0.1°C per second (20x)
4. from 50°C, decrease by 1°C, hold for 2minutes, ramp at 0.1°C per second (10x)
5. from 40°C, hold for 10minutes, ramp at 0.1°C per second

III. Removal of unbound oligos

Unbound oligos from gDNA were removed using 25µl SPB per each well. Two wash steps using SW1 and a third wash step using 60% ethanol were performed to ensure complete removal of unbound oligos.

IV. Extension and ligation of bound oligos

Products containing the targeted regions of interest flanked by sequences required for amplification were the product of a DNA polymerase extension from the hybridized upstream oligo followed by ligation to the 5' end of the hybridized downstream oligo using a DNA ligase (Figure 7).
Summarily, 22 µl ELB/ELE mixture were added to each well and then the plate was run on a thermal cycler at 37°C for 45min, 70°C for 20min and hold at 4°C.

V. Amplification of libraries

A unique combination of two index adapters were added to the extension-ligation products of each tumour DNA by PCR which were required for further cluster formation as well as barcoding of individual tumour sequencing reads (Figure 7). Succinctly, 4µl of each index i7, 4µl of each index i5 and 20µl EDP/EMM mixture were added to each well. Then, considering the appropriate number of PCR cycles (for 385-700 amplicons it is 28 cycles and for FFPE samples it is added 1 more cycle, resulting in 29 PCR cycles total), the PCR program was performed as follows:

Initial denaturation	95 °C	3 min	
Denaturation	98 °C	20 sec	29x
Annealing	67 °C	20 sec	
Extension	72 °C	40 sec	
Final extension	72 °C	1 min	
Hold	10 °C	∞	

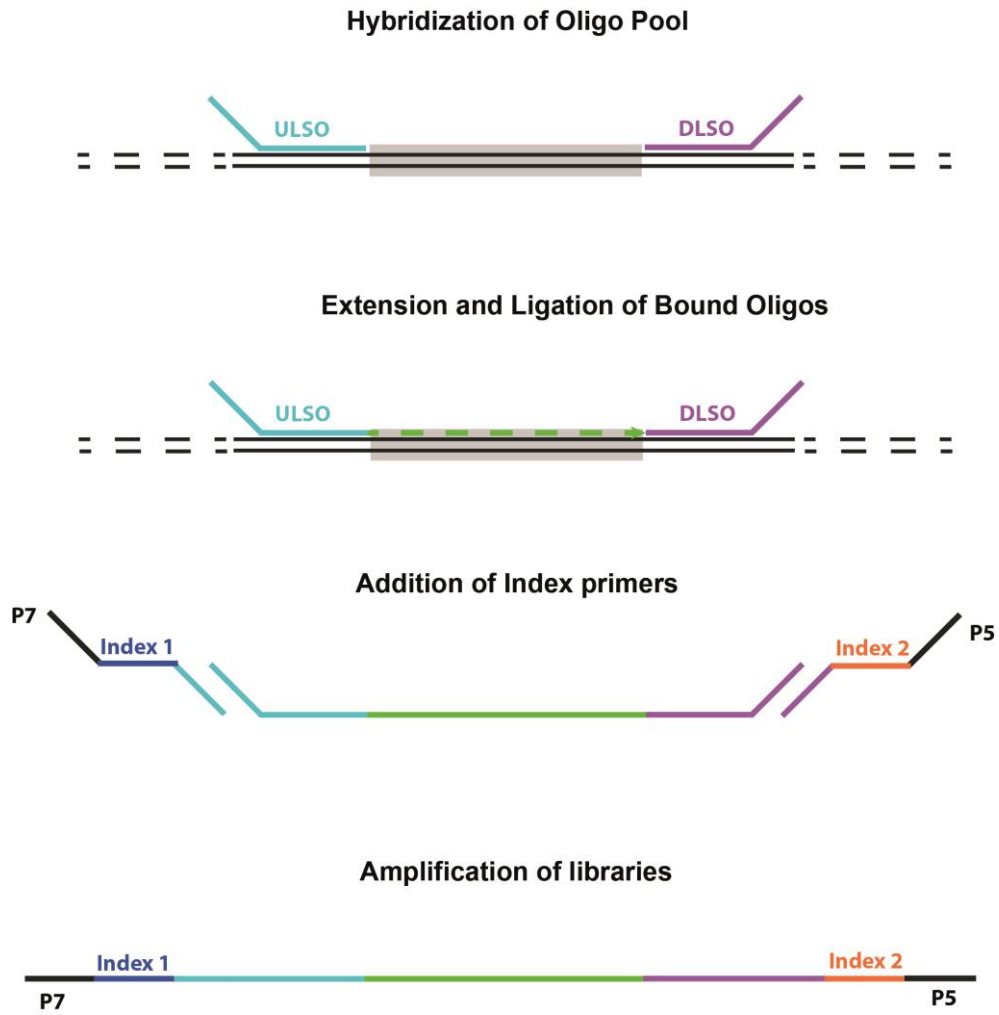


Figure 7 - NGS library preparation overview
The first step of the library preparation for NGS is the hybridization of a custom oligo pool containing upstream and downstream oligos specific to the target regions of interest (Upstream Locus-Specific Oligo – ULSO - and Upstream Locus-Specific Oligo - DLSO), crucial for further amplification. Then, the hybridized ULSO and DLSO are connected by DNA polymerase extension from the ULSO through the targeted region followed by ligation to the 5' end of the DLSO using a DNA ligase, leading to the formation of products containing the targeted regions of interest flanked by the sequences needed for amplification. Index adapter oligos (A5XX and A7XX) from Illumina containing a unique sample index and P5/P7 adapters, which are required for the template to be compatible with the flow cell for cluster formation, are added. At this point, extension-ligation products containing its singular barcode are ready to proceed for their amplification. (figure adapted from Illumina® TruSeq® Custom Amplicon Library Preparation Guide)

VI. Purification of libraries

For the purifying the PCR products 36µl SPB and 25µl RSB per well were used.

VII. Assessment of the libraries prepared

In order to assess the success of libraries' preparation, 1 µl of each purified library was run on an Agilent Bioanalyzer (Figure 8) using a DNA 1000 chip according to *Agilent DNA 1000 Assay Protocol* (Agilent Technologies, Waldbronn). The bioanalyzer traces' outputs were analysed using 2100 Expert software (Agilent Technologies, Santa Clara) in which the PCR products for 200 bp amplicons were expected to be approximately at 350bp size trace.

Figure 8 - Agilent bioanalyzer to assess NGS library preparation
Agilent bioanalyzer containing the DNA 1000 chip loaded with glioma samples and DNA ladder to assess the libraries' quality

VIII. Normalization of libraries

For an effective sequencing process, it is crucial that each library has the same representation in pooled libraries. Briefly, 45 µl LNA1/LNB1 was added to each library well and washed with 45µl LNW1 for two times. 30µl of fresh 0.1N NaOH was then added to each well and 30 µl of its supernatant was transferred from the LNP plate to the corresponding well in the SGP plate which also contained 30 µl LNS2.

IX. Pool libraries

Equal volumes of normalized library (5µl) were combined in a single tube. Then, 12µl of combined libraries were diluted in 588µl hybridization buffer and heat denatured. Pooled libraries were loaded into the thawed MiSeq reagent cartridge for the sequencing run.

3.2.1.3 Sequencing

Sequencing of the previously prepared libraries was performed as indicated in the *MiSeq System User Guide* and it is schematically shown in Figure 9.

The previously prepared adapter-ligated library fragments were loaded into a flow cell which contained oligonucleotides complementary to Illumina adapter sequences leading to the hybridization of the fragments to the flow cell surface. Then, through bridge amplification, each bound fragment was amplified generating clonal clusters. Illumina's sequencing technology employs the sequencing-by-synthesis approach in which the nucleotides are fluorescently labelled and act as reversible terminators of the amplification reaction. Then, in each cycle, the four nucleotide analogues are simultaneously added and incorporated by the DNA polymerase; these nucleotides are chemically blocked (replacement of the 3'-OH group for a 3'-o-azidomethyl group) to prevent the incorporation of more than one nucleotide in each cycle. Once the nucleotide is incorporated, the flow cell is imaged and the fluorescence emission wavelength and intensity recorded from each cluster are used to identify the base incorporated. The nucleotides that have not been used are washed away and the chemical blockade of the 3' end is removed with tris-(2-carboxyethyl)-phosphine, to continue the synthesis of the chain. Thus, once the fluorescence signal is collected, a new cycle begins, repeating the process, each base separately, until the sequencing of each fragment is finished (72,74,65).

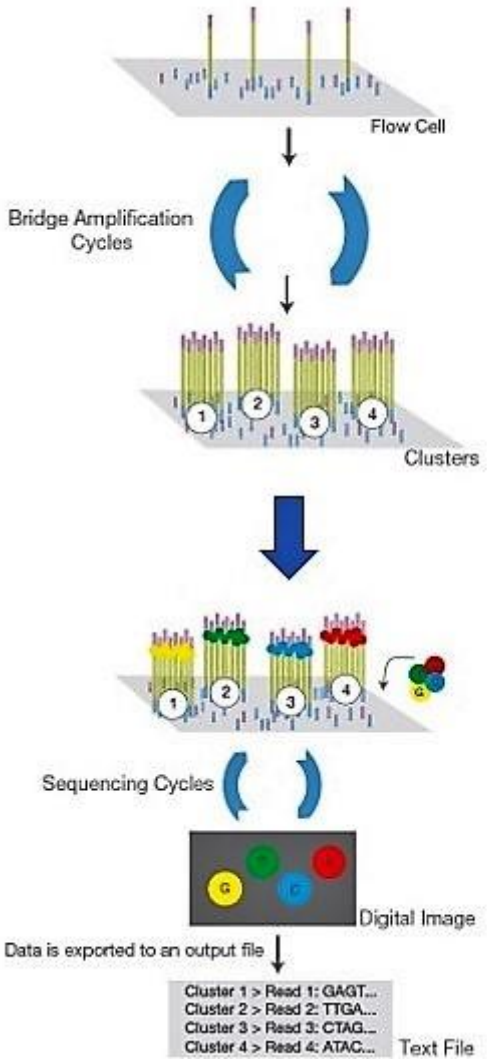


Figure 9 - Representation of Illumina NGS cluster amplification and sequencing
Single-stranded, adapter-ligated fragments are hybridized to primers immobilized on the flow cell surface; the free end of a hybridized fragment "bridges" to a complementary oligo on the flow cell surface; multiple cycles of denaturation and extension result in localised cluster amplification (bridge amplification) generating clusters of DNA with identical molecules. Sequencing reagents, such as DNA polymerase and all four fluorescently labelled nucleotides are added simultaneously to the flow cell channels for incorporation into the cluster fragments; the flow cell is imaged and the emission data recorded from each cluster is used to identify the incorporated base; data is exported to an output file. (figure adapted from (72))

3.2.1.4 Data analysis

Global data of the NGS run including sequence coverage and quality parameters were evaluated using Illumina Sequencing Analysis Viewer 1.8.37 (Illumina Inc, Germany). More particular sequencing outputs were analysed resorting Illumina VariantStudio Variant Analysis Software 2.2 (Illumina Inc, Germany) and Integrative Genomics Viewer (IGV) 2.3 (Broad Institute, USA). These last two bioinformatics software packages aligned the reads to the human reference genome GRCh37 (hg19) so the differences between the reference and the sequenced amplicons were easily identified (Figure 10). This alignment was facilitated due to the manifest file, provided by Illumina, containing the coordinates of the DNA fragments in the library, allowing the direct comparison of the sequencing reads to the respective regions of the reference genome and, in this way, the software does not have to look for the reads in the entire genome.

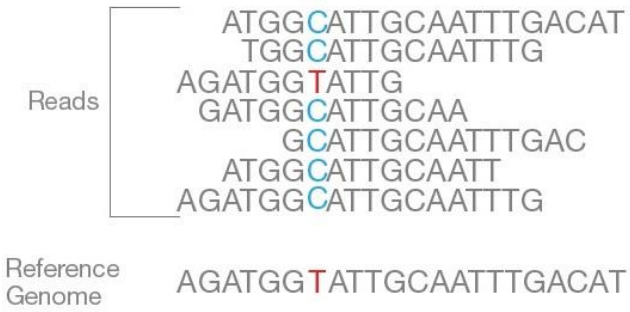


Figure 10 - NGS data analysis methodology
Illumina VariantStudio Variant Analysis Software 2.2 (Illumina Inc, Germany) explanatory image comparing NGS assay reads to reference genome in which a single-nucleotide variation T>C is observed. (figure adapted from (72))

- Filtering process

As NGS technology provides thousands of variations for each sample, it is crucial to apply extra-filtering techniques to taper our research and so to minimize false positive findings.

Illumina Sequencing Analysis Viewer (SAV) furnishes critical quality metrics generated by the MiSeq real-time analysis software, such as the percentage of reads identified – "pass-filter" (PF). This parameter, PF, is the percentage of reads which pass chastity filter, the internal quality filtering process of Illumina sequencers. Chastity, according to Illumina, is defined as *"the ratio of the brightest base intensity divided by the sum of the brightest and second brightest base intensities"*; so, in the first 25 cycles only one base

call can have a chastity value below 0,6 to allow clusters pass filter (75). This first step of filtering clears away the least trustworthy clusters from the image analysis outputs.

Then, using Illumina VariantStudio Variant Analysis Software 2.2, three different grades of filtering were applied (Table 11).

In order to select the most reliable new-mutations and to further confirm some by Sanger sequencing, two major criteria were contemplated: technical issues (is the mutation true or false?) and biological consequence (is it damaging or benign?). To the first selection criteria, besides a variation frequency >20% (the higher the better), only read depths above 100 were considered. To the second selection criteria, only frameshift variants and deleterious/damaging missense variants were selected. In addition, visualization of the mutation on the IGV 2.3 (Integrated Genome Viewer) was performed to assess if the mutation seemed to be true or not: the software allows not only to see individual sequencing traces (forward and reverse), so one would notice if a mutation artefact only occurs in one sequencing direction, but also to see other sequence artefacts up- or downstream of a particular variant, which are both indicators of fixation artefacts.

Table 11 - Filtering process performed during NGS data analysis
The filtering process, which originated three different filtering grades, was carried concerning 4 major filtering settings: (i) MiSeq instrument pass filters, which consists in an internal quality filtering process of the sequencer; (ii) population frequency of the detected variants, which was always set to be under 5% not to create e.g. population polymorphisms bias; (iii) DNA variation frequency, which firstly was set to be higher than 10% and then higher than 20% to further restrict the outputs and (iv) biological consequences, which at first were all shown (intronic and exonic mutations) and then were confined only to mutations inside genes and further excluded splicing mutations. As such, the most stringent filter used, "very very stringent", included missense, frameshift and stop gained/lost variations and inframe insertions and deletions with a variation frequency higher than 20%.

Filtering process	**Stringent**	**Very stringent**	**Very very stringent**
MiSeq instrument	Pass filters (e.g. strand bias)		
Population frequency	< 5%		
Variation frequency	>10 %	> 20%	
Biological consequences	All	Only inside genes	Only inside genes; No splicing mutations

3.2.2 Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction was performed for the amplification of DNA segments of interest.

Oligonucleotide primers (with approximately 20 nucleotides) are designed complementary to the DNA region of interest. PCR thermocycling causes denaturation of double-stranded DNA into single-stranded DNA allowing the primer to anneal the region of interest and the *Taq* DNA polymerase to mediate the elongation.

This methodology leads to exponential *in vitro* reproduction of a nucleic acid fragment and it is therefore the choice for a quick amplification of a DNA segment of interest which is crucial for its analysis.

In this work, PCR was performed to amplify *IDH1*, *BRAF* and *H3F3A* hot spot mutation sites (codon 132, codon 600 and codons 27 and 34, respectively) and several other mutation sites according to the NGS results. PCR products were further used for pyrosequencing or Sanger sequencing validation, correspondingly.

3.2.2.1 PCR reaction setup

The standard PCR reaction mixture consisted as described below:

Reagent	**Amount**
gDNA (25 ng/µl)	3,0 µl
Master mix 2x	12,5 µl
CoralLoad concentrate 10x	2,5 µl
Forward primer 10pmol/µl	1,5 µl
Reverse primer 10pmol/µl	1,5 µl
Water	4,0 µl
Total volume:	25 µl

A positive-control (mutated DNA), a negative-control (WT DNA) and a blank (H_2O) were also performed to control and validate the assay.

3.2.2.2 PCR program

The typical PCR thermocycling was performed as follows:

1. Initial denaturation	95 ºC	15 min	
2. Denaturation	94 ºC	30 sec	50x
3. Annealing	60 ºC	30 sec	
4. Extension	72 ºC	30 sec	
5. Final extension	72 ºC	10 min	
6. Hold	10 ºC	∞	

In specific cases in which the PCR product was too low or non-existent, an optimization of the procedure was achieved by adjusting the annealing temperature to 54ºC as well as the elongation time to 40sec.

3.2.2.3 PCR quality control: agarose-gel electrophoresis

Agarose-gel electrophoresis was performed to survey PCR amplification – verification of the quality and size of PCR products.

Therefore, 3µl PCR-products were first used for control on a 3% agarose-gel with pUC19 as length standard (34-501 bp). Typical 3% agarose-gels were prepared by dissolving 3g peqGOLD agarose in 100ml 1xTBE buffer (from 10xTBE composed by 540g tris base, 275g boric acid, 200ml 0,5M EDTA pH 8 and water until 5l solution) and adding 8µl Midori Green. The gels were run for 10min at a voltage of 150V in BioRad Sub-Cell® GT Cell chamber.

DNA fragments with intercalated Midori Green were visualized by UV illumination at 254nm. The image of the agarose-gel was captured using Quantum ST4 system and analysed using Quantum-Capt software image enhancement.

3.2.3 Pyrosequencing

This method was used to validate NGS results of almost all giant-cell glioblastoma samples (n = 22) for hotspot mutation in codon 132 of *IDH1*, in codon 600 of *BRAF* and in codons 27 and 34 of *H3F3A*. Two samples were not analysed by pyrosequencing because of lack of DNA material.

Pyrosequencing is a technology which is based on sequence detection providing a rapid and accurate quantification of sequence variation (76). Sequencing by this methodology is achieved by a synthetic process in which nucleotides are dispensed one at a time and the reading is carried out as each of the four nucleotides are incorporated into the template strand replication (Figure 11) (65,77).

Firstly, the DNA segment of interest is amplified by PCR using biotinylated primers so that the strand to serve as the pyrosequencing template is biotinylated too. The single-stranded biotinylated template anneals with the pyrosequencing prime and this complex is incubated with the enzyme-mix composed by DNA polymerase, ATP sulfurylase, luciferase and apyrase, as well as the substrate-mix composed by adenosine 5' phosphosulfate (APS) and luciferin. When the first dNTP is added to the reaction, if it is complementary to the base in the template strand, DNA polymerase catalyzes its addition to the sequencing primer. releasing a pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide. In the presence of APS, PPi is converted to ATP by ATP sulfurylase Then, using ATP as cofactor, luciferase mediates the conversion of luciferin to oxyluciferin generating visible light in proportional amounts to the amount of ATP. This light is detected by a charge coupled device camera and seen as a peak (pyrogram) whose height is proportional to the amount of nucleotides incorporated. In parallel, the unincorporated nucleotides and ATP are continuously degraded by apyrase to avoid these residues interfering in later cycles and enabling further dNTP additions. Ultimately, the complementary DNA strand is completed and the nucleotide sequence is determined from the set of signal peaks in the pyrogram trace (76,65,77).

Many different assays can be executed simultaneously and the duration of each assay is exclusively dependent on the number of dispensations needed to cover the region of interest (77).

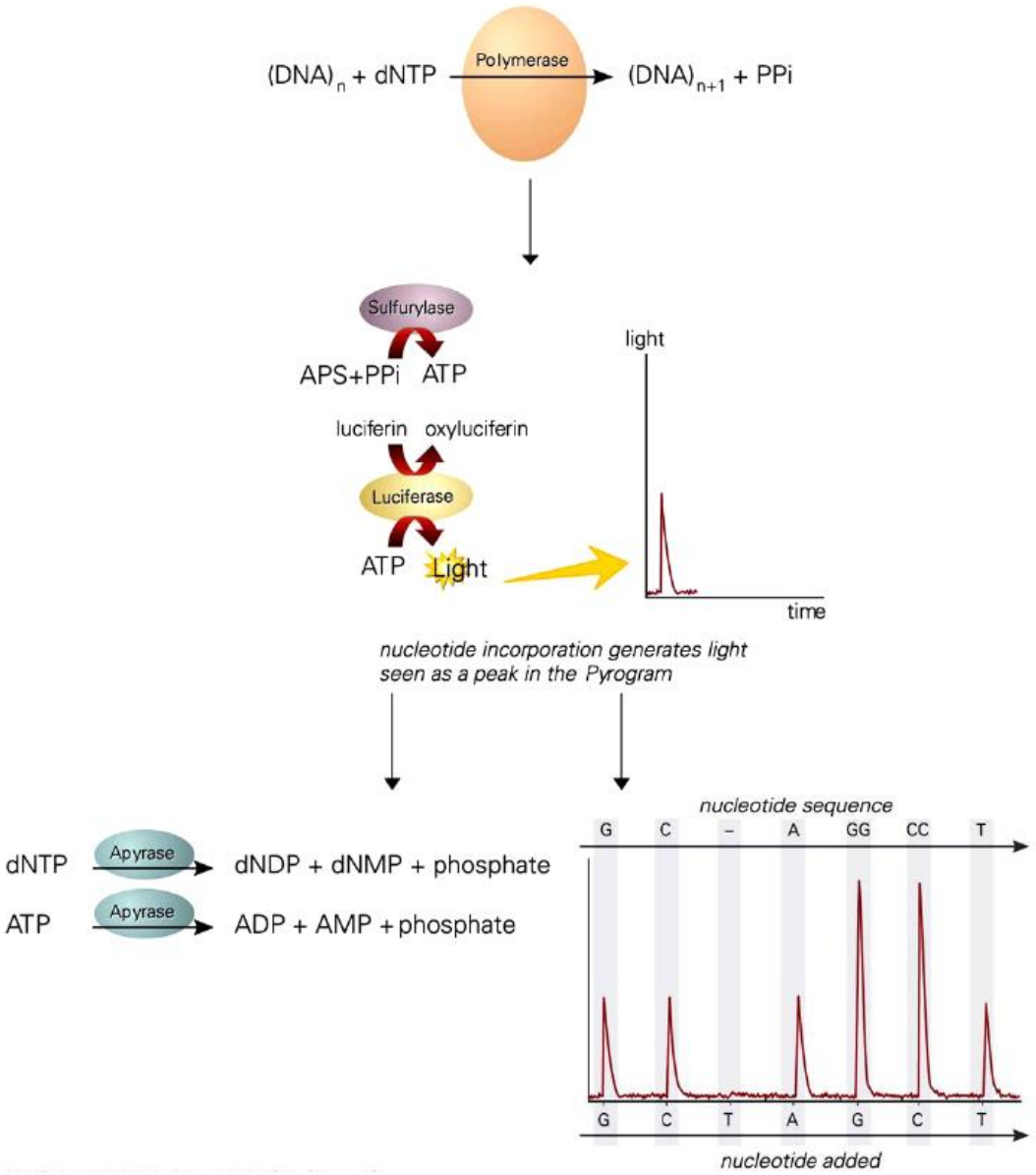


Figure 11 - Pyrosequencing chemistry
DNA polymerase catalyzes the addition of a dNTP complementary to the template strand releasing a PPi in each incorporation. In the presence of APS, PPi is converted to ATP by ATP sulfurylase. Luciferase uses ATP as cofactor for the conversion of luciferin to oxyluciferin releasing light in the process. This light is detected by a charge-coupled device camera and seen a peak (pyrogram) whose height is proportional to the amount of dNTP incorporated. (figure origin: (78))

3.2.3.1 Pyrosequencing reaction setup and program

Pyrosequencing of the PCR products was performed using pyrosequencing primers and the PyroMarkGold Reagents. 10µl of each PCR product was pipetted on a 8-strip pyrosequencing tubes in addition to 30µl mastermix composed of 18,5 µl binding buffer, 1,5µl streptavidin-sepharose beads and 10µl HPLC-H_2O. The samples were shaken for 10 minutes on the microplate shaker (1400 rpm). Subsequently, 1µl of pyrosequencing

primers mixed with 24µl of annealing buffer was pipetted into a fresh pyrosequencing primer plate. The purification of the DNA-bead samples for sequencing was preformed resorting a vacuum prep tool as described below:

Vacuum ON		
1	HPLC-H_2O	5 sec
2	HPLC-H_2O	5 sec
3	Samples (beads uniformly sucked)	10 sec
4	70% EtOH	5 sec
5	Denaturation solution	5 sec
6	Wash-buffer	10 sec
Vacuum OFF		
7	Primer plate (beads released → annealing)	3 min
8	Incubate 80ºC (denaturation)	2 min
Vacuum ON		
9	HPLC-H_2O	5 sec
10	HPLC-H_2O	5 sec

Afterwards, the pyrosequencing plate was cooled down and placed in the PyroMark Q24 instrument (Qiagen, Hilden). Furthermore, the pyrosequencing cartridge containing PyroMark enzyme-mix, substrate-mix and dNTPs was prepared according to the sample sheet previously elaborated using the PyroMark Q24 Software (Qiagen, Hilden) and placed in the sequencing instrument. At last, the USB flash drive containing the sample sheet (with the assays to perform and correspondent dispensation orders) is opened on the sequencing instrument enabling the sequence process to begin.

3.2.3.2 Data analysis

Pyrogram outputs were analysed using the PyroMark Q24 Software (Qiagen, Hilden). All samples were compared with the positive and negative controls of each assay.

3.2.4 Sanger sequencing

Sanger sequencing was performed to validate 103 non-hotspot mutations selected from the NGS outputs. In order to assess these 103 selected mutations, primers were designed resorting UCSC Genome Browser (University of California, USA).

Sanger sequencing, also known as the chain-termination method, is the gold-standard DNA sequencing technology and it is the process of selective incorporation of chain-terminating dideoxynucleotides (ddNTPs) by DNA polymerase during in vitro DNA replication (79,80,81).

A single-stranded DNA template, a DNA polymerase, a DNA primer, normal nucleotides (dNTPs) and modified nucleotides (ddNTPs) which stop DNA strand elongation are required to perform this method (81).

During the sequencing, DNA polymerase incorporates nucleotides to a growing chain which are selected by base-pair matching to the single-stranded DNA template. DNA polymerase adds not only dNTPs but also its analogues, ddNTPs. As the growth occurs by the formation of a phosphodiester bridge between the 3'-OH group on the primer and the 5'-phosphate group of the incoming nucleotide, the extension of the DNA terminates when the ddNTPs are added due to these modified nucleotides lacking the 3'-OH group that is required for the ligation between two nucleotides. Therefore, chain elongation is terminated selectively at A, C, G or T and as each of the four ddNTPs are fluorescently-labelled, the sequencing is read while each of the different fluorescent signs are detected during capillary electrophoresis, resulting in the final sequence scheme – chromatogram (Figure 12) (79,80,82).

3.2.4.1 Sanger sequencing reaction setup

This method was not performed in the lab. PCR products and respective designed primers were sent to Eurofins company to execute the assay. Thus, on each well of a prepaid barcoded 96-well plate (Eurofins Genomics, Germany), 10µl of PCR product and 5µl of water were added. Each PCR product was analysed twice – forward and reverse sequencing. Precisely identified, 15µL of each diluted designed primer (10pmol/µL) were added in 2ml tubes with safe-lock. Then, the prepared material was sent to the company in a pre-paid pre-addressed shipping envelope (Eurofins Genomics, Germany).

3.2.4.2 Data analysis

Sequencing results were sent by the company and analysed resorting BLAT search from UCSC Genome Browser (University of California, USA) and further compared with the respective previous results from NGS.

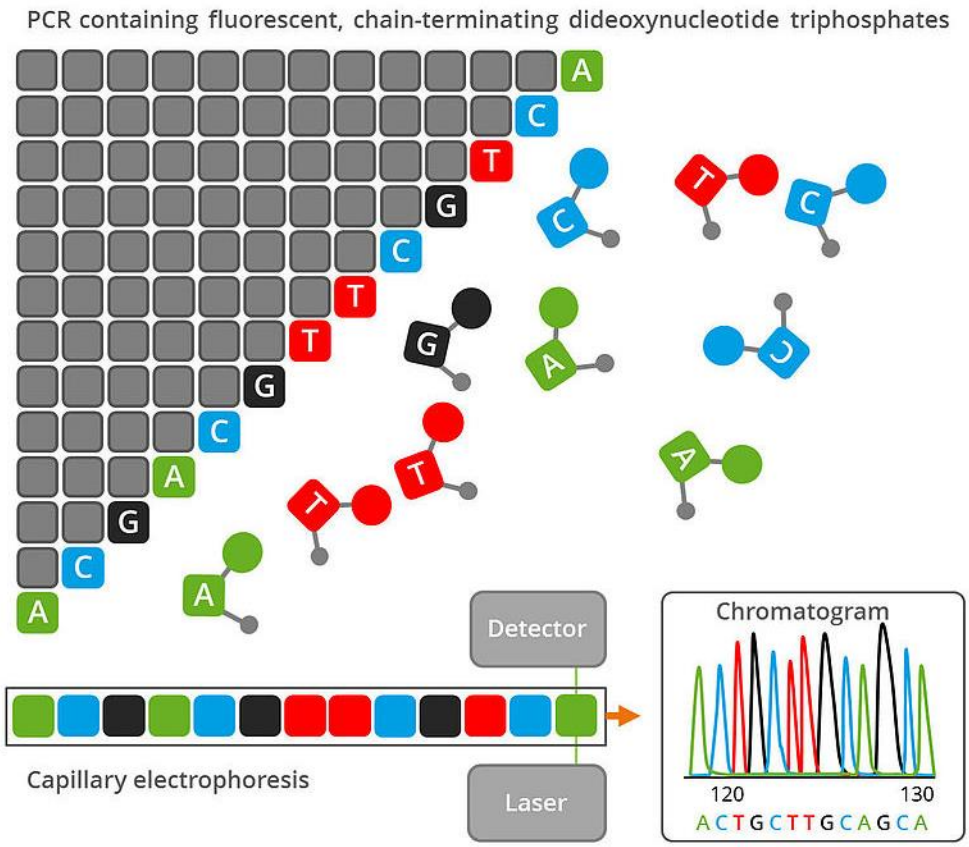


Figure 12 - Sanger sequencing technology
Gold-standard DNA sequencing method based on the detection of fluorescently-labelled chain-terminating nucleotides (ddNTPs) that are incorporated by a DNA polymerase during the replication of a template. (figure origin: (83))

4 RESULTS

4.1 Assessment of NGS libraries preparation

At the end of each Bioanalyzer run, it was created a virtual agarose-gel by its software (Figure 13) allowing to assess the NGS library preparation's success for each sample. With this, the analysis of DNA libraries' quality was facilitated: the best libraries were scored of 3 points (strong band), good libraries of 2 (satisfactory band), poor libraries of 1 (weak band) and the not visible libraries (no band) of 0. 90% of the libraries showed a successful preparation process (only 3 libraries didn't show any band).

1) Real-time bioanalyzer trace

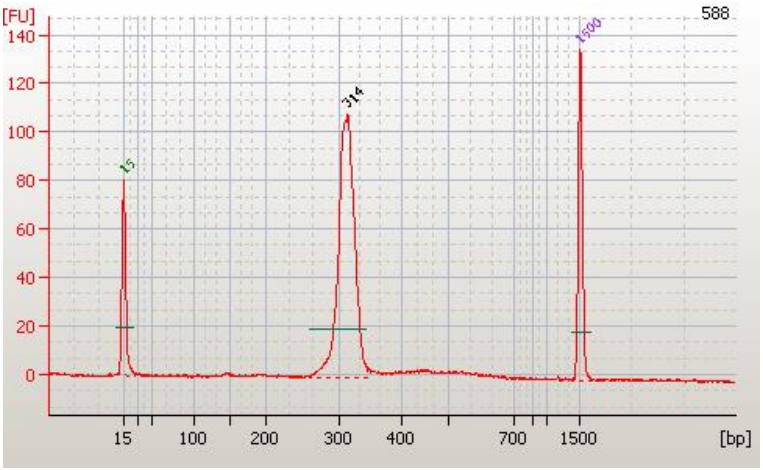
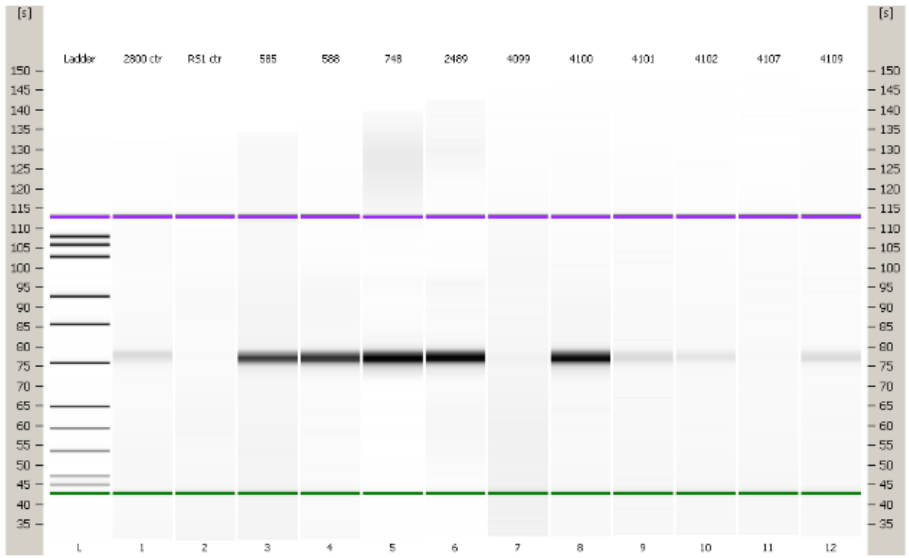
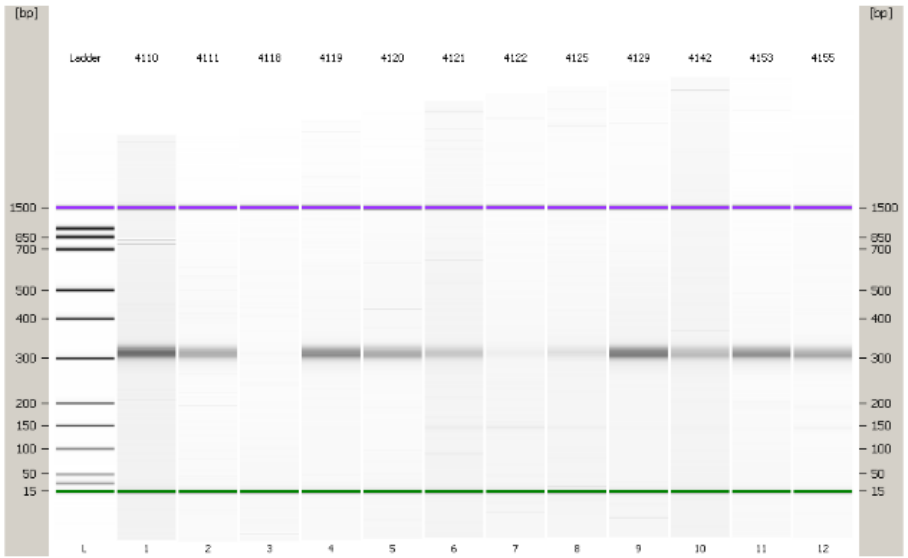


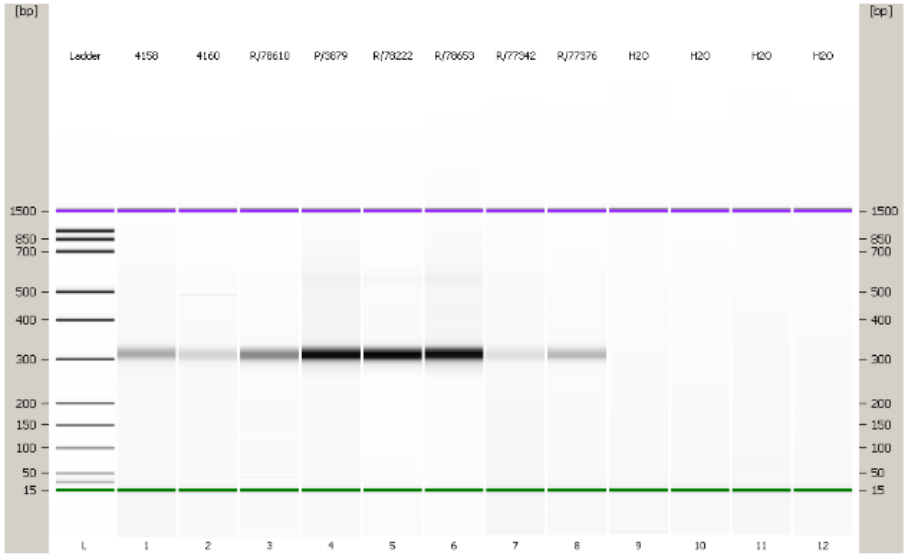
Figure 13 - Agilent bioanalyzer results for NGS libraries' quality
1) Real-time bioanalyzer trace representing a very good library with the expected PCR product size ~310bp and **2)** Virtual agarose-gels performed by Agilent bioanalyzer software where it is possible to assess the NGS library preparation's success: at the first two wells are represented the control samples 2800 (positive control) and RS1 (negative control); further, samples 1-4, 6 and 26-28 were scored of 3 (strong band); samples 11, 12, 14-16, 19-25 and 30 were scored of 2 (satisfactory band); samples 7, 8, 10, 17, 18 and 29 were scored of 1 (weak band); samples 5, 9 and 13 were scored of 0 (no band).

Figure 13 - continued

2) Virtual agarose-gels performed by Agilent bioanalyzer software







4.2 Next-Generation Sequencing: overall analysis

The targeted gene panel was appraised in twenty-four giant-cell glioblastoma DNA samples (12 from adult patients and 12 from infant patients) and in six other glioma DNA samples with previously known molecular diagnostics; all the DNA samples came from tumour tissues surgically removed and preserved through FFPE procedure.

DNA library preparation and sequencing were prosperous in 30 of the 30 samples (100%), despite quality differences.

Illumina Sequencing Analysis Viewer 1.8.37 (Illumina Inc., Germany) showed a total number of reads of 19.878.072 of which approximately 95,1% were successfully mapped - reads identified (PF) - with a 115 bp mean read length. As quality indicator it was obtained a median Phred quality score of 28 (Q=27,75).

A huge diversity of genetic alterations was pinpointed including single-nucleotide variations, which were the most representative (~86%; C>T more frequent), followed by deletions (~8%) and insertions (~6%). Exonic variations were found more commonly (~66%) than non-exonic variations (~34%).

ATRX, *TP53*, *CIC* and *NF1* were the most commonly mutated genes among all samples. One of the least mutated gene, and with less variations found, was *TERT* (including its promotor), which hot spot mutations 228C>T and 250C>T were detected only in four samples (three of the first type and one of the second). The genes with more variants outputs were *NF1*, *ATRX* and *EGFR*.

No sample had zero important genetic variations in NGS outputs.

4.3 Impact of the filtering process on NGS data analysis

NGS data analysis showed an average of 979 variants for each sample when no filter had been applied and an average of 28 variants when the most stringent filter had been applied. A similar rate (35 no filter : 1 most stringent filter) was observed when evaluating the filtering process for each gene.

In this manner, only roughly 3% of NGS variants outputs were available to further data analysis procedures.

The impact of the filtering process for each DNA sample on NGS data analysis is represented below (Figure 14).

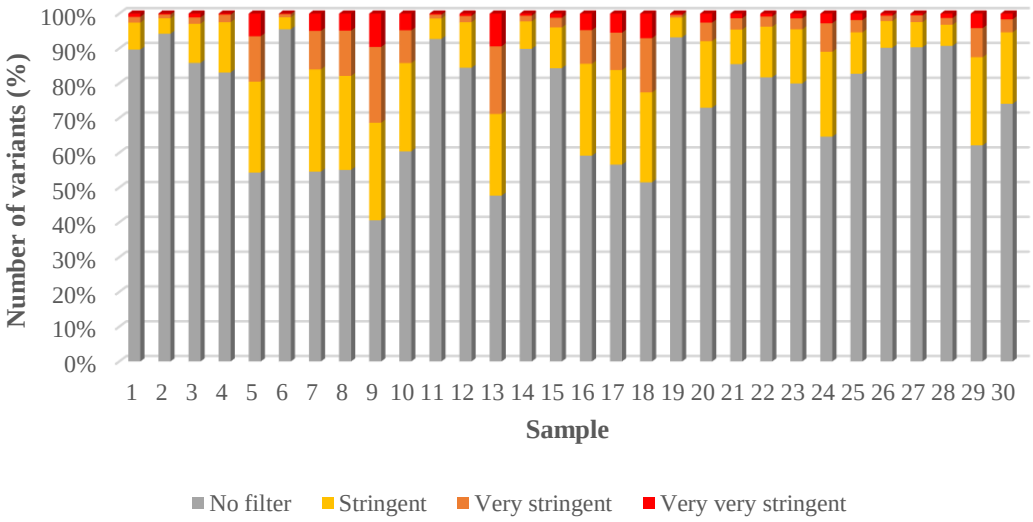


Figure 14 - Impact of the filtering process for each sample during NGS data analysis
The more stringent the filtering process got, the less number of variants remained intact to further analysis use. The filtering process was capable of reducing in approximately 97% the number of variants found in the NGS run, leaving only 3% reliable to further data analysis procedures.

4.4 Correlation between quality of the DNA libraries and NGS outputs

To determine if there was any correlation between quality of the DNA libraries and the NGS outputs, the first parameter was compared with the most significant NGS technological parameters: (i) the percentage of reads identified by Illumina Sequencing Analysis Viewer and (ii) the number of variants detected.

4.4.1 DNA libraries' quality and percentage of reads identified (PF)

As it can be observed in Figure 15, DNA libraries with the best quality result in the Bioanalyzer evaluation were found to have a higher percentage of reads identified in SAV as well as DNA libraries with the worst quality result had lower percentage of reads identified. These two parameters showed to carry a statistically significant ($p = 1{,}34E\text{-}07$) very strong positive correlation ($r = 0{,}8$).

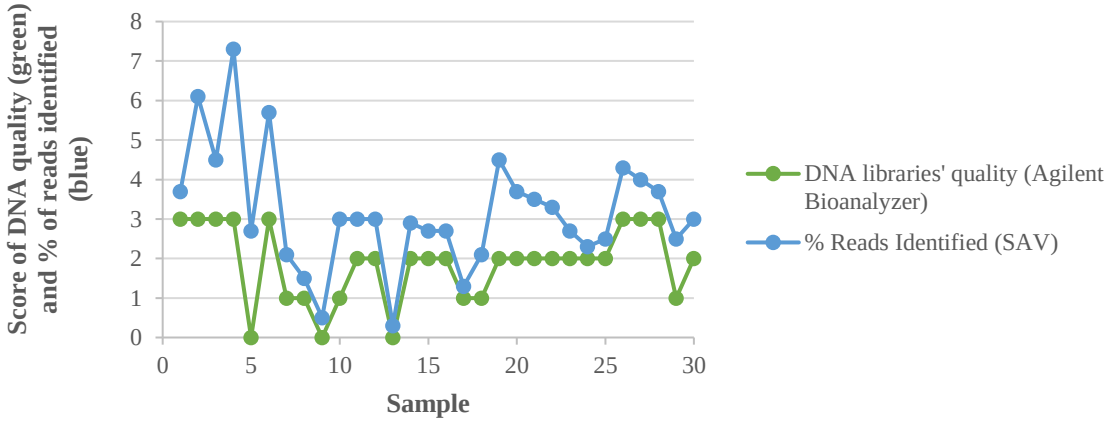


Figure 15 - DNA library's quality of each sample and correspondent percentage of reads identified
Dispersion plot showing a relationship between DNA libraries' quality obtained from Agilent Bioanalyzer results (scored of 0 to 3, from the poorest to the greatest library quality) and the percentage of reads identified for each sample in Illumina Sequencing Analysis Viewer (varying from 0 to 8%).

4.4.2 DNA libraries' quality and number of variants detected

The total number of variants detected in NGS (no filter) didn't show any significant dependence with the sample DNA's quality ($r = 0{,}2$; $p = 0{,}2$). However, when the most stringent filter was applied, a moderate negative correlation was observed: samples with better DNA quality had generally less variants detected ($r = -0{,}63$; $p = 0{,}0002$).

4.5 Sensitivity and specificity of NGS technology

4.5.1 Detection of previously identified mutations

NGS accurately detected 6 of 6 (100%) previously known hotspot mutations which had been identified by pyrosequencing. All of them were well-known damaging SNVs, each one from a different glioma sample.

Also, functioning as negative control, NGS didn't spot 7 of 7 (100%) specific hotspot mutations that had been previously assessed as wild-type by pyrosequencing analysis: two in *IDH1* codon 132, two in *IDH2* codon 172, one in *H3F3A* codon 27, one in *H3F3A* codon 34 and one in *BRAF* codon 600.

4.5.2 Validation of NGS hotspot mutations results for codon 132 of *IDH1*, codon 600 of *BRAF* and codons 27 and 34 of *H3F3A* with pyrosequencing

2 of 2 samples (100%) *IDH1* positive to hotspot mutation in codon 132 in NGS analysis were validated by pyrosequencing. NGS results showed also *BRAF* V600E hotspot mutation in five DNA samples which all five (100%) were afterwards corroborated by pyrosequencing analysis. Additionally, four mutations in the *H3F3A* gene (three K27M and one G34R) were detected by NGS and could be confirmed by pyrosequencing (100%).

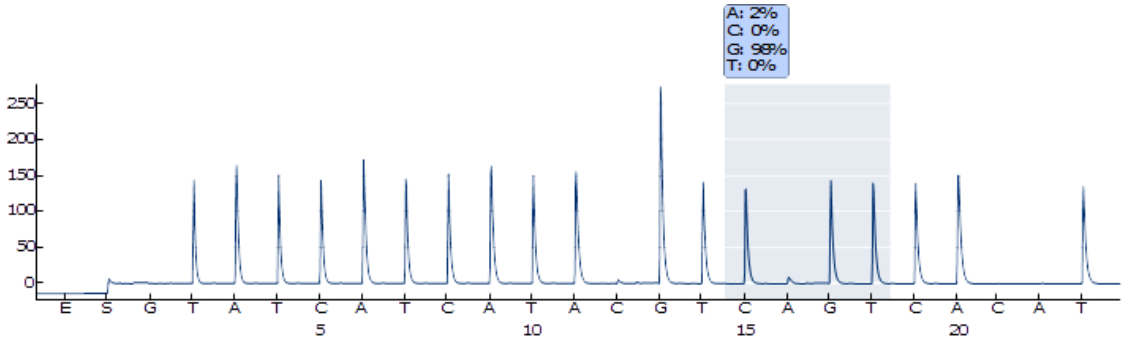
The same success rate was achieved validating the wild-type NGS results for the first two genes: 20 of 20 samples (100%) negatives to *IDH1* hotspot mutation in codon 132 and 17 of 17 samples (100%) negatives for *BRAF* hotspot mutation in codon 600 present in NGS outputs were all confirmed by pyrosequencing analysis.

All these results are schematically represented in Table 12, Figure 16 and Figure 17.

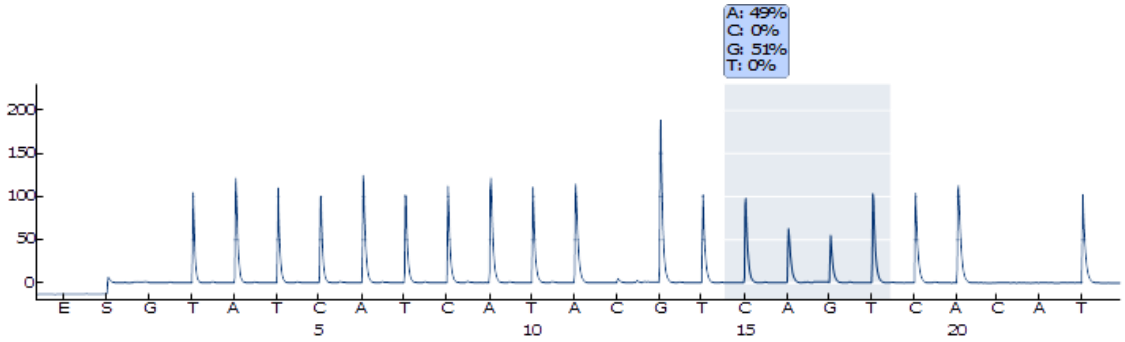
Table 12 - Detection of previously identified mutations and validation of NGS hotspot mutations
This representative table exhibits two technologies' results which were compared with each other - DNA hotspot mutations detected first by pyrosequencing and then compared to the NGS outputs obtained (green) and GcGBM DNA hotspot mutations detected first by NGS and afterwards validated by pyrosequencing (blue). The correlation between the frequencies of the DNA variation detected by both technologies was found to be strong (r = 0,7) and statistically significant (p=0,0034).

Gene	Coordinate	Variant, biological consequence and protein position/ amino acid changes	Type of DNA variation	Frequency of the variation in pyrosequencing analysis	Frequency of the variation in NGS outputs
H3F3A	Chr1:226252135	A>A/T; Missense; K27M	SNVs	31%	34,68%
H3F3A	Chr1:26252156	G>G/T; Missense; G34V		63%	65,69%
H3F3A	Chr1:226252155	G>G/A; Missense; G34R		26%	41,38%
HIST1H3B	Chr6:26032206	A>A/T; Missense; K27M		20%	25,66%
IDH1	Chr2:209113113	C>C/T; Missense; R132C		40%	38,43%
TERT promotor	Chr5:1295228	Upstream gene variant 228C>T		21%	43,50%
IDH1	Chr2: 209113112	C>C/T; Missense; R132H		43%	40,77%
				47%	47,47%
BRAF	Chr7: 140453136	T>T/A; Missense; V600E		30%	30,66%
				30%	27,11%
				24%	20,77%
				20%	20,74%
				38%	38,27%
H3F3A	Chr1: 226252135	A>A/T; Missense; K27M		40%	28,39%
				45%	37,89%
				48%	29,35%
	Chr1: 226252155	G>G/A; Missense; G34R		29%	36,52%
p = 0,0034					

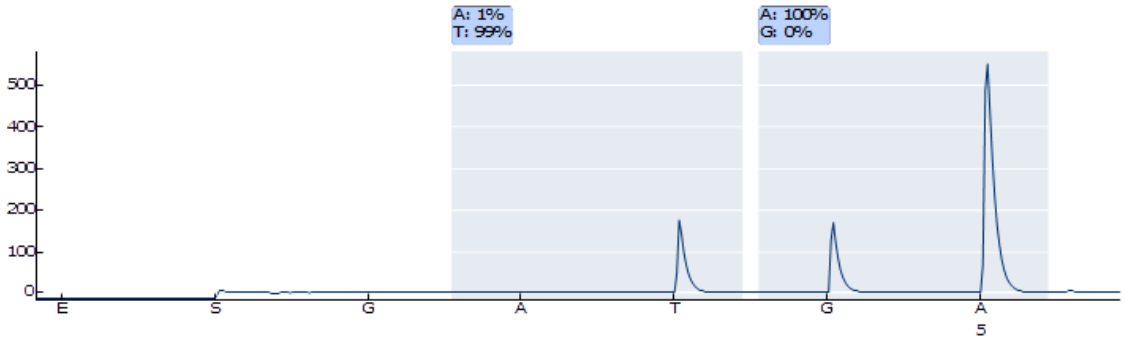
(A) *IDH1* Wild-type



(B) *IDH1* Mutant R132H



(C) *BRAF* Wild-type



(D) *BRAF* Mutant V600E

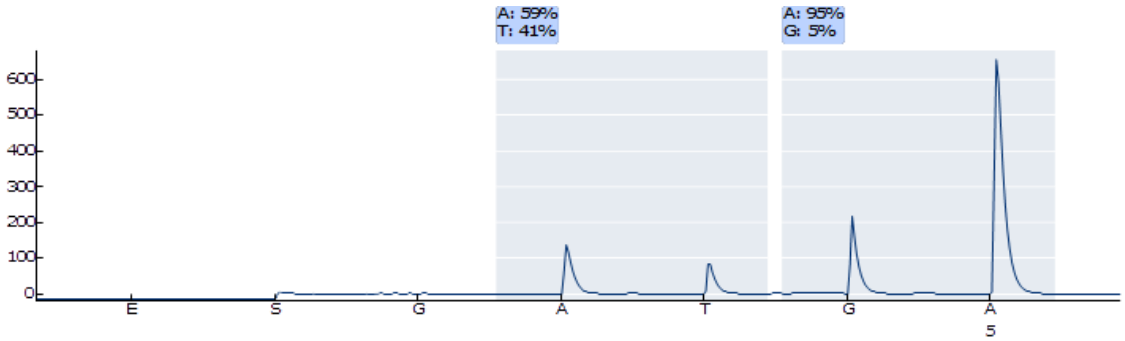
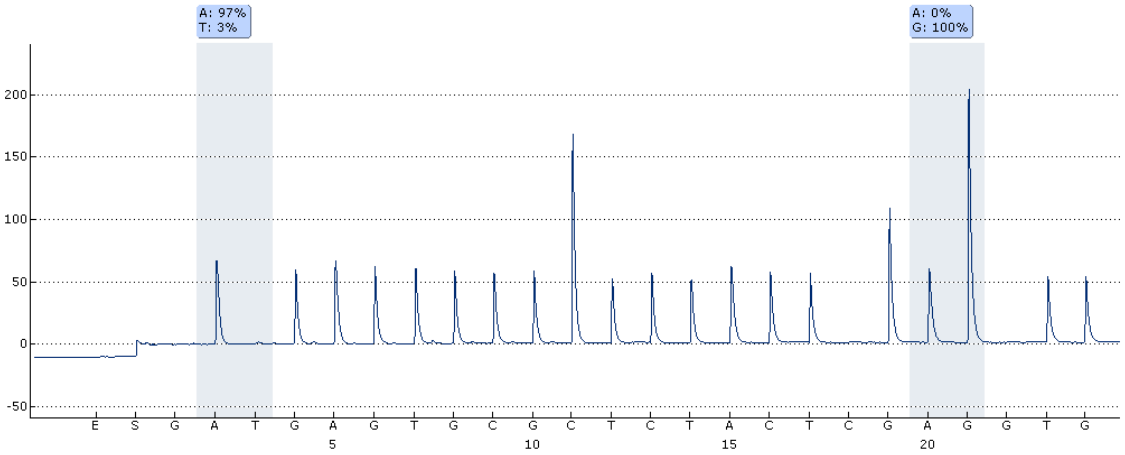


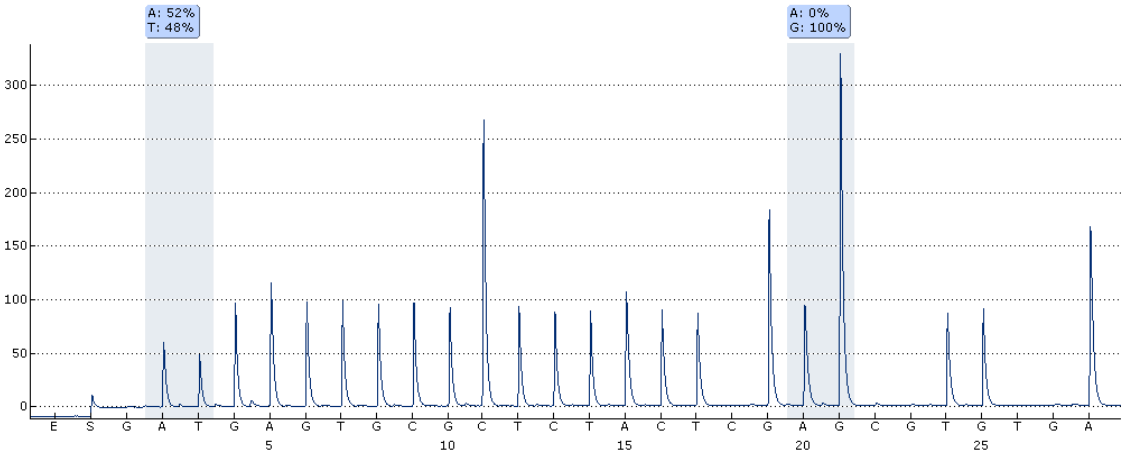
Figure 16 - Pyrogram representations of each NGS result validated for hotspot mutations in codon 132 of *IDH1*, in codon 600 of *BRAF* and in codons 27 and 34 of *H3F3A*
(A) *IDH1* wild-type; **(B)** typical *IDH1* mutant pyrogram where G>G/A, resulting in the substitution of arginine to histidine (R132H, 49%); **(C)** *BRAF* wild-type; **(D)** typical *BRAF* mutant pyrogram where T>T/A, resulting in the substitution of valine to glutamic acid (V600E, 59%); **(E)** *H3F3A* wild-type; typical H3F3A mutant pyrograms where **(F)** A>A/T, resulting in the substitution of lysine to methionine (K27M, 48%) or **(G)** G>G/A, resulting in the substitution of glycine to arginine (G34R, 20%)

Figure 16 - continued

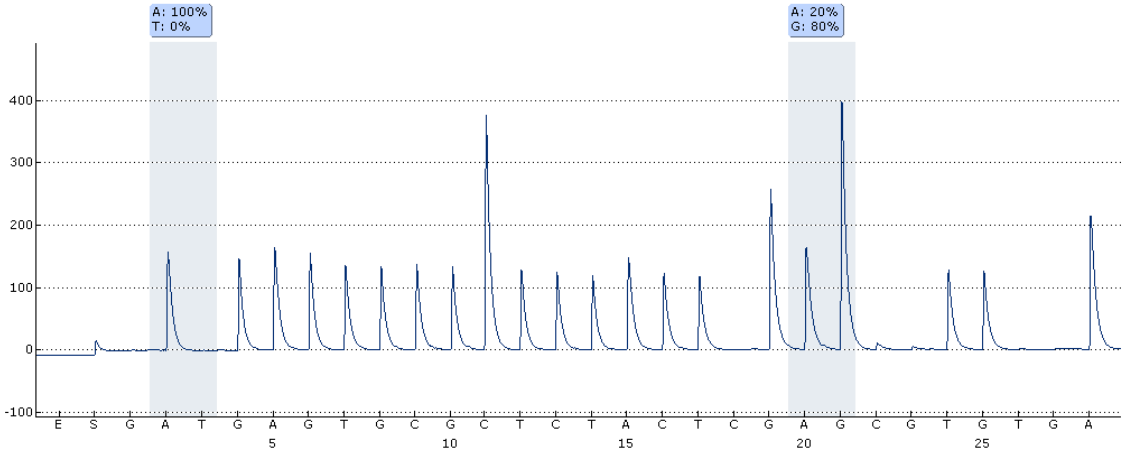
(E) *H3F3A* Wild-type



(F) *H3F3A* Mutant K27M



(G) *H3F3A* Mutant G34R



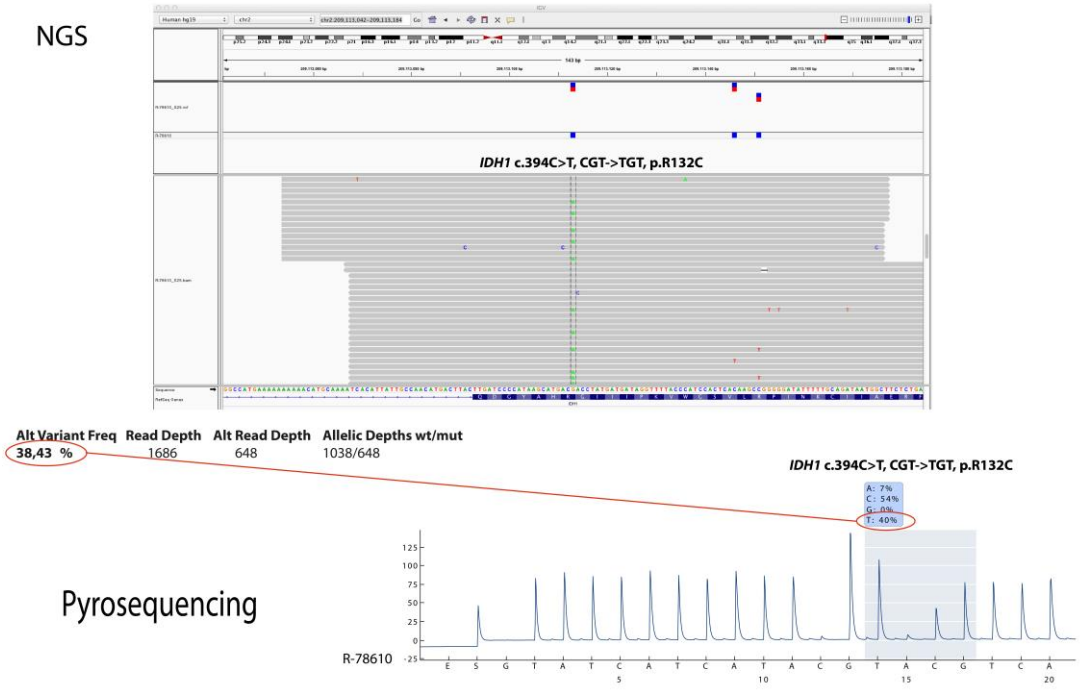


Figure 17 - Comparison of an established hot spot mutation (*IDH1* R132C) identified by NGS and pyrosequencing
This figure shows the IGV image of a region of the *IDH1* gene containing the R132C well-known mutation (upper part of the figure) together with its data from Illumina software containing technical settings (variant frequency, read depth) as well as the pyrogram of this hot spot mutation. Both technologies safely identify the same mutation of *IDH1* with comparable variant frequencies (38,43% NGS; 40% pyrosequencing).

4.5.3 Validation of NGS novel mutations results with Sanger sequencing

At first, only 2 of 22 (~9%) new mutations selected from NGS outputs were confirmed by Sanger sequencing. Among the genes investigated, and the number of variations analysed for each one, were: *BRAF* (one), *CIC* (five), *EGFR* (one), *FUBP1* (five), *HIST1H3B* (three), *H3F3A* (one), *IDH1* (three), *PIK3R1* (two) and *PPM1D* (one). These first confirmed mutations were one *PIK3R1* deletion CTCAGTT>CTCAGTT/C (inframe deletion at Chr5:67589598) with 97% variation frequency in NGS and one *CIC* insertion G>G/GC (frameshift variant at Chr19:42795885) with 96% variation frequency in NGS.

For the second set of Sanger's validations, 81 more novel mutations identified by NGS were selected. 2 samples did not show a PCR product for Sanger sequencing and as such they couldn't be further analysed. 32 of 79 (40,5%) new mutations analysed were confirmed by Sanger sequencing. Among the genes investigated, and the number of variations analysed for each one, were: *ACVR1* (five), *ATRX* (nineteen), *CIC* (four), *EGFR* (three), *ERBB2* (one), *H3F3A* (one), *NF1* (fourteen), *PIK3CA* (two), *PIK3R1*

(three), *PPM1D* (one), *PTEN* (four) and *TP53* (twenty-two). In this second set of confirmations, a richer pattern of confirmed mutations was found: 25 SNVs, 6 deletions and 1 insertion, with variant frequencies varying from 24,3% to 100%. A SNV found by NGS and further validated by Sanger sequencing is shown in Figure 18.

Hence, a total of 103 new variants detected by NGS were selected and 101 (98%) could be further analysed by Sanger sequencing which in turn validated 34 of 101 (~34%) new mutations.

Almost all false positive results (58,2%) were alterations C>C/T with a variation frequency in NGS between 21% and 93% and alterations G>G/A with a variation frequency in NGS between 22% and 66%.

Some of the analysed genes had several new mutations next to each other in a sample, so that with just one pair of primers it was possible to assess more than one mutation. But all of these cases came out to be negative for all mutations.

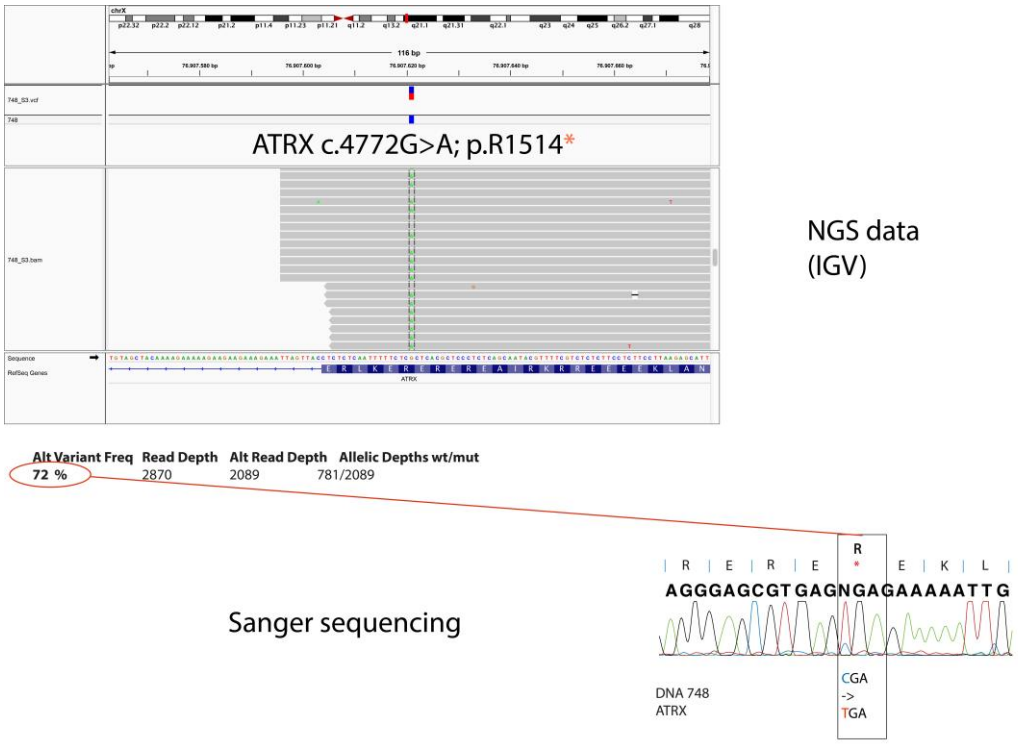


Figure 18 - Novel mutation identified by NGS and its Sanger sequencing further validation
This figure shows a verified NGS result (upper part of the figure) together with the respective region analysed by Sanger sequencing (lower part of the figure). Both technologies identified a *ATRX* c.4772G>A; pR1515* mutation with similar frequency. This stop gain variation leads to a premature termination of the protein and can be considered deleterious.

4.6 Extensive analysis of NGS outputs concerning the DNA variants landscape

4.6.1 Most and least commonly mutated genes

As it is shown in Figure 19, the most commonly mutated genes, transversely to all samples, were *ATRX*, *TP53*, *CIC* and *NF1*; the least frequently mutated genes were *IDH2*, *FGFR1*, *TERT* (including its promotor) and *HIST1H3B*.

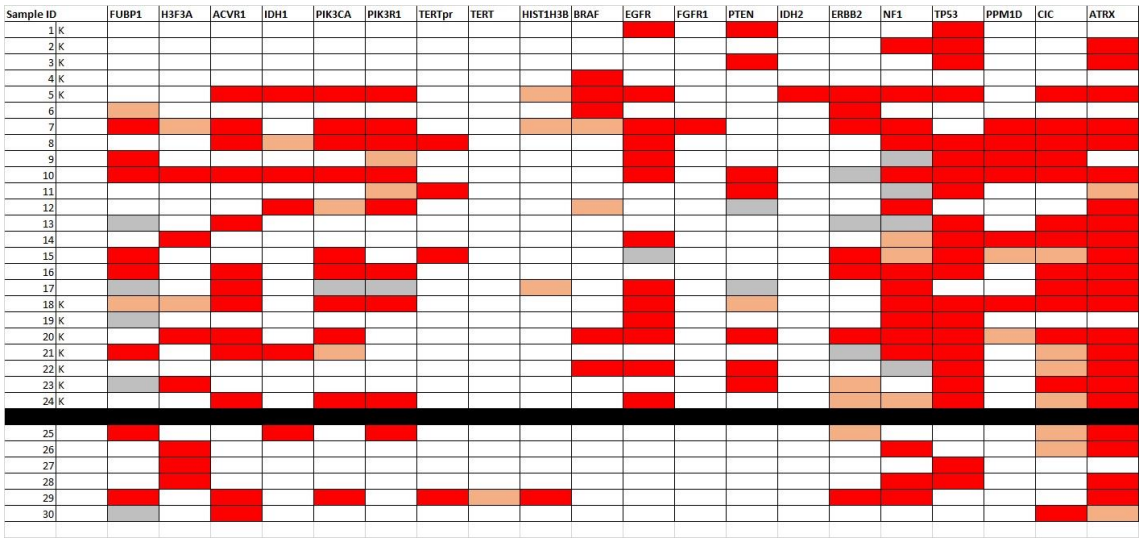


Figure 19 - Expeditious view of mutated genes per sample
Global data analysis summary of genetic outputs for each sample. In red the genes mutated for sure; in orange the genes which are probably mutated but it's possibly not damaging; in grey the results technically uncertain (lower read depths); in blank the wild-type (not mutated). "K" stands for child sample.

4.6.1.1 GcGBM distinctive mutational profiles in adults and children

A comparison between GcGBM outputs and the other control gliomas outputs was not performed. The only reason for inclusion of the six glioma samples was the fact that information on hot spot mutations were available from those cases allowing them to function as controls, oppositely to GcGBM whose DNA abnormalities were collected for the first time. Also, the difference in the samples' proportion of each (24 of GcGBM vs 6 control gliomas) would lead to biased results and conclusions.

As such, looking only at GcGBM outputs, different mutated gene patterns were easily checked. For instance, *TERT* promotor hotspot mutations were only found in adult's samples.

FUBP1 mutations were mostly detected in adult's samples (83%) as well as *PPM1D* mutations (83%), *PIK3R1* mutations (75%), and *CIC* mutations (67%).

Oppositely, *BRAF* mutations had a higher incidence in children samples (80%) such as *PTEN* mutations (71%).

All the other genes had a similar rate of mutants analysing adult and children samples.

4.6.2 Disparities in the number of variants detected concerning different genes

The genes with more variants outputs were *NF1*, *ATRX* and *EGFR*; the smallest number of variants was detected in the *TERT* gene (including its promotor), *HIST1H3B* and *IDH2* (Figure 21).

A moderate correlation was observed between the number of variants detected and the gene size (p = 0,002) (Figure 20).

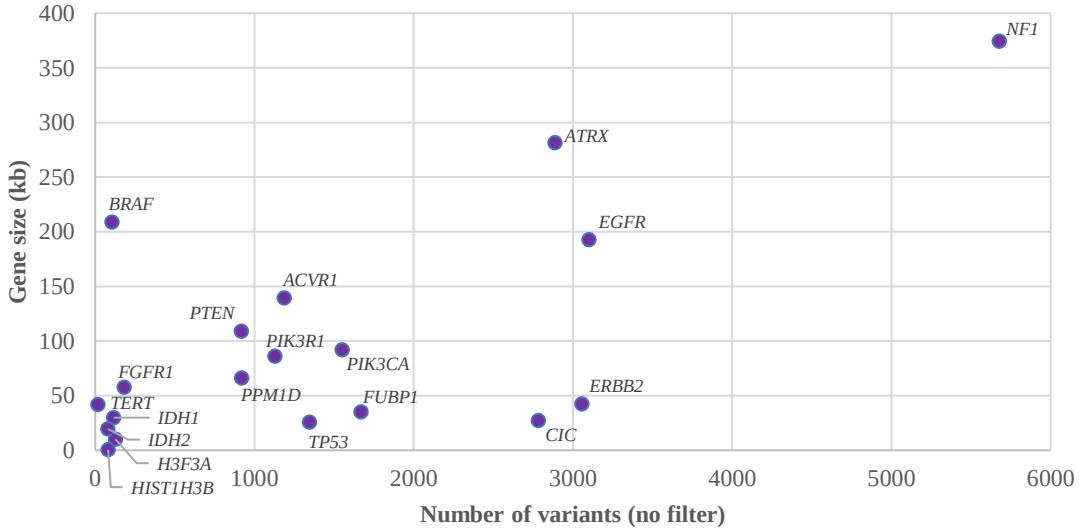
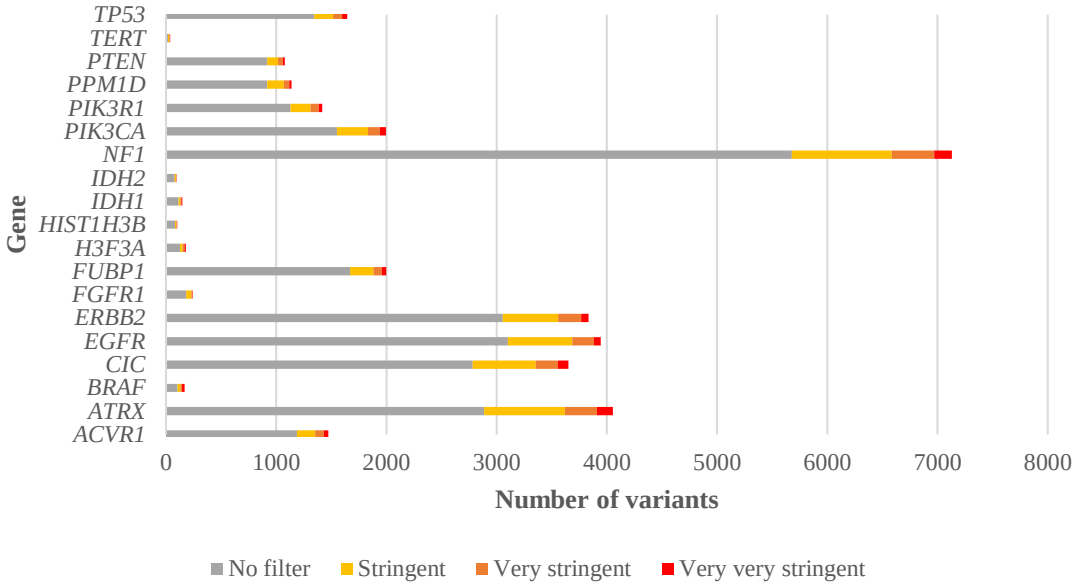


Figure 20 - Relation between the gene size and the number of variants detected
A moderate correlation (r = 0,7) was found concerning the gene size and the correspondent number of variants detected. Each circle represents a gene and, as it is shown in this graph, when the gene size was higher, often a larger number of variants was detected (p = 0,002).

I. Number of variants detected *vs* gene analysed – representation of the difference in the number of variants detected for each gene



II. Number of variants (%) *vs* gene analysed – representation of the proportion of variants detected for each gene after different filtering grades were applied

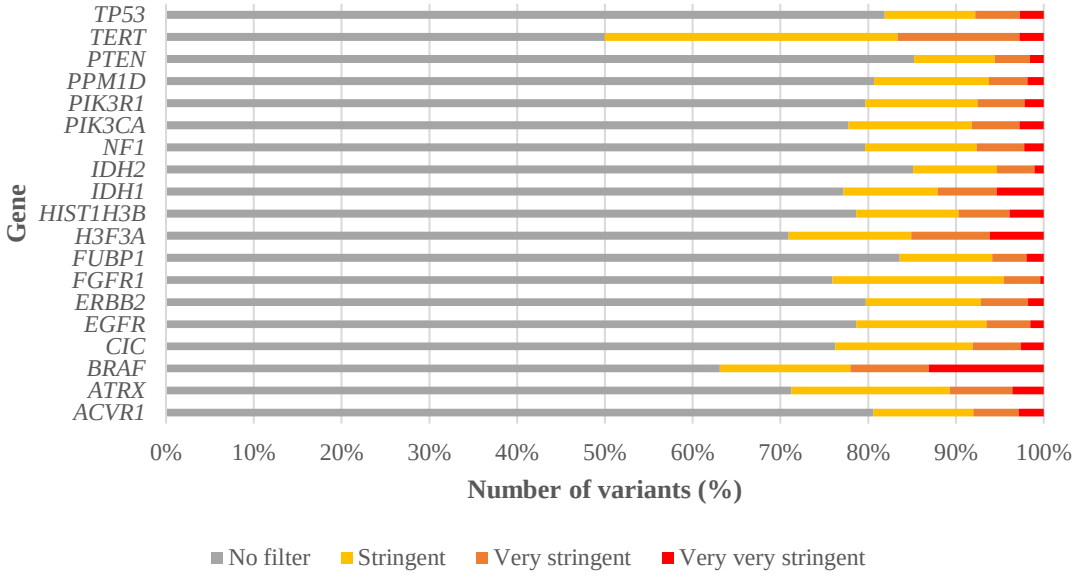
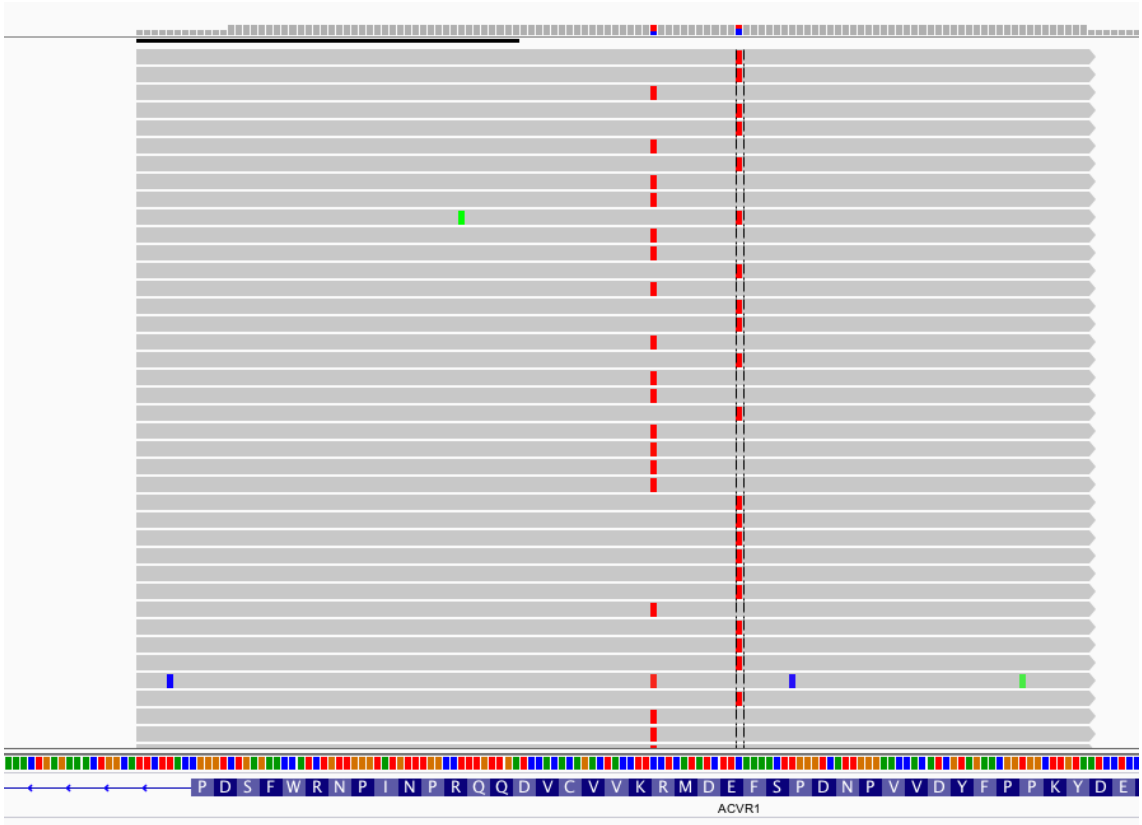


Figure 21 - Number of variants detected per each gene
I. in this graph it is easily observable the colossal difference between each gene's variants detected. *NF1*, *ATRX* and *EGFR* were the genes with more variants detected when no filter was applied; however, *CIC*'s number of variations detected overcomes EGFR's when the most stringent filter was applied. On the contrary, smaller numbers of variants were detected in *TERT* gene (including its promotor), *HIST1H3B* and *IDH2*, both when no filter and when most stringent filter were applied. **II.** this graph displays the different proportion of variants detected for each gene when consecutive filtering grades were applied. *BRAF* was the gene with higher percentage of variants filtered (20,8%) and *FGFR1*, oppositely, was the gene with the shorter proportion of variants filtered (0,6%), comparing to PF outputs. Nonetheless, the filtered variants consisted, in general, in 1-9% of PF outputs (average 3%).

4.6.3 Multiple mutation sites in the same gene

Even after applying the most stringent filter, there were cases in which a sample had the same gene with multiple important mutations detected (stop gains, damaging SNVs, deletions and insertions). These cases were mostly, but not only, found concerning the genes with more altogether variants found (*NF1*, *ATRX*, *EGFR* and *CIC*), that is, in the genes with higher size. However, all these cases which were analysed by Sanger sequencing came out to be negative for all mutations. An illustrative example of these findings is displayed in Figure 22.

a) *ACVR1*



b) *NF1*

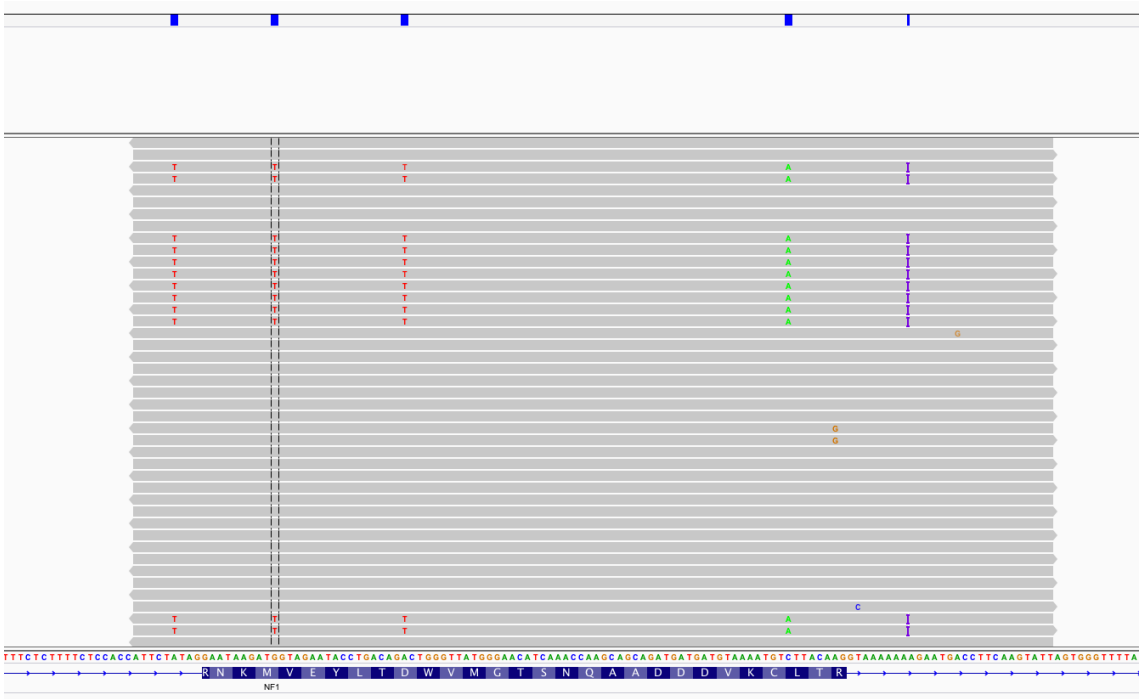


Figure 22 - Illustration of multiple mutations detected in *ACVR1* gene and in *NF1* gene
These IGV images show two examples of simultaneous detection of several mutations close to each other in a region of the *ACVR1* gene **a)** and in a region of the *NF1* gene **b)**. These variants turned out to be false positives which point that these findings can act as an indicator of false positive data (e.g. fixation artefacts C>T G>A transitions)

5 DISCUSSION

The classification of gliomas into subtypes based not only on histological features but also molecular markers has come a long way (16). Last year, the 2007 WHO classification of brain tumours was revised and changed considerably due to the increased demand of molecular analysis for establishing a more accurate diagnostics of gliomas which has originated the 2016 WHO new classification already including genetic markers in the glioma diagnostic pathway, in addition to the histological features traditionally considered (42,14). Currently, diagnostic neuropathology laboratories apply individual testing for selected biomarkers, such as mutations in *IDH1/2*, *TERT*, *BRAF*, *ATRX* and *H3F3A* as well as 1p/19q codeletion (42). These individual assessments are performed resorting to immunohistochemistry with mutation-specific antibodies against *IDH1*-R132H, *BRAF*-V600E and *H3*-K27M, fluorescence in situ hybridisation (FISH) for detection of 1p/19q codeletion and tumour DNA sequencing through classical Sanger sequencing or pyrosequencing for detection of individually selected mutations (42,84,85,86,87). However, novel genetic abnormalities are continuously arising (42) and as more research into the different genetic features among gliomas is performed, further insight into the diagnostic, prognostic and therapeutic value will be acquired (47). Hereby, the application and validation of a new diagnostic method capable of performing robust molecular profiling in brain tumours is increasingly required in order to improve their management (42,64,63). Recent studies pointed out that NGS-based approaches making use of commercial or customized gene panels fulfil this need and so may become a valuable tool in future glioma diagnostics providing a cost and time efficient alternative to currently used conventional sequencing techniques (42,41,67).

In this work a NGS customized gene panel for diagnostic sequencing of 19 genes frequently aberrant in gliomas was successfully established and verified.

The process of DNA preparation and sequencing was found to operate greatly, holding a very good quality score. With roughly one week of preparation of 30 libraries and three days of sequencing them and obtaining the outputs, hundreds of genetic variants for each sample were collected. This contrasts with conventional technologies, such as pyrosequencing and Sanger sequencing in which each sequencing reaction evaluates only one gene position for each sample (67,69).

The filtering proceedings were impetuous concerning the data analysis, reducing in approximately 97% the number of variants found in the NGS run, remaining 3% reliable to further data analysis procedures. However, this 3% still represent a great amount of data, which not all are real or clinically important. In this way, and as mentioned in previous studies (64,67,73), it is crucial to point the huge impact of the filtering process when NGS technology is used and so the need of further studies to continue the discovery of the best track to do it so the remain variants are worth taking in account for being clinically relevant to the diagnostics of brain tumours.

In this work, it was observed the importance of the library preparation concerning the NGS outputs. Only 10% of the libraries didn't have a successful quality result in the Bioanalyzer assessment, even though they showed sequencing results on NGS. As it was displayed before, DNA libraries with the worst quality result had lower percentage of reads identified in the NGS technology. So, the same way that in other sequencing technologies better quality libraries ease and accurate the sequencing process (88), the NGS results lead us to a similar conclusion that as the DNA quality rises, the easier is to this technology to sequence the DNA. Also, considering the number of DNA sequence changes detected in NGS, no significant correlation was found between this parameter and the DNA quality when no filter had been applied but when the most stringent filter was applied, samples with better DNA quality had generally less variants detected. These findings point, once again, that the filtering process is imperative to the NGS data analysis carrying a different impact on the results of a good or a bad library, restricting more effectively the outputs of the better than of the worse. With this, better DNA quality libraries present more reliable results showing less DNA sequence changes.

TruSeq NGS method demonstrated to hold an immense sensitivity and specificity for detection of diagnostically relevant DNA variants such as hotspot mutations in *IDH1* (R132), *H3F3A* (K27;G34), *HIST1H3B* (K27), *BRAF* (V600) and *TERT* promotor (C228), showing a high concordance with the single-gene approach results previously or further performed. In the same sample presenting an *IDH1* mutation previously identified by pyrosequencing, the 1p19q codeletion biomarker had been also previously detected and, curiously, for this sample, NGS showed an insertion in *CIC* gene (96% variation frequency) later confirmed by Sanger sequencing. This fact is consistent with formerly reported data in which authors conclude that there is a high correlation between *CIC* mutation and 1p19q codeletion (89,90). Even though the current version of the applied

glioma panel does not allow a quantitative assessment of chromosomal arms 1p and 19q, the presence of *CIC* mutation and *FUBP1* mutation together with an *IDH1* mutation clearly confirms the diagnostics of an oligodendroglioma for this tumour specimen (14).

NGS also identified several novel mutations in the 19 genes collection. In the first set of new-mutations analysed by Sanger sequencing, only 9% could be confirmed. 95% of the false-positives were SNVs C:G > T:A, which were probably formalin-induced artefacts (cytosine deamination to uracil, frequently found in ancient DNA) – Figure 23 -, showing the difficulty of NGS to identify FFPE tissue artefacts which is in line with previous studies (91,92). In this manner, no reliable conclusion could be pointed because these mutations represented 86% of the novel mutations selected to assess. As such, a larger second set of confirmations was performed in order to eliminate the bias that resulted from selecting bad DNA/NGS data on the first selection of variants. In this second set, the results improved substantially and one could validate 41% of the selected new mutations. Thus, a total of 101 selected new-mutations detected by NGS were further analysed by Sanger sequencing which has validated 34 of them (34%). Nevertheless, most of the selected mutations from NGS outputs came out to be negative and SNVs C:G > T:A still represented the majority of the novel mutations selected and 58,2% of the false-positive DNA variant calls in NGS analysis. Currently, there are not enough facts about the veracity of this kind of variants and in order to correctly evaluate NGS capability to detect real novel mutations, two approaches should be further studied: (i) performing a treatment of the FFPE-DNA with uracil–DNA glycosylase (91); (ii) choosing other type of novel mutations detected by NGS to further confirm with Sanger sequencing, such as deletions, insertions and SNVs avoiding C:G > T:A.

All the results presented on the extensive analysis of NGS outputs concerning the genes' variants landscape demand deeper studies. Nonetheless, in general, the most commonly mutated genes detected in our diversified samples and the distinctive mutational profiles found in adults and children GcGBMs were consistent with the overall gliomas' molecular characterizations displayed in Table 1. However, and once again, further validation and a deeper research is crucial to allow stronger and more reliable conclusions.

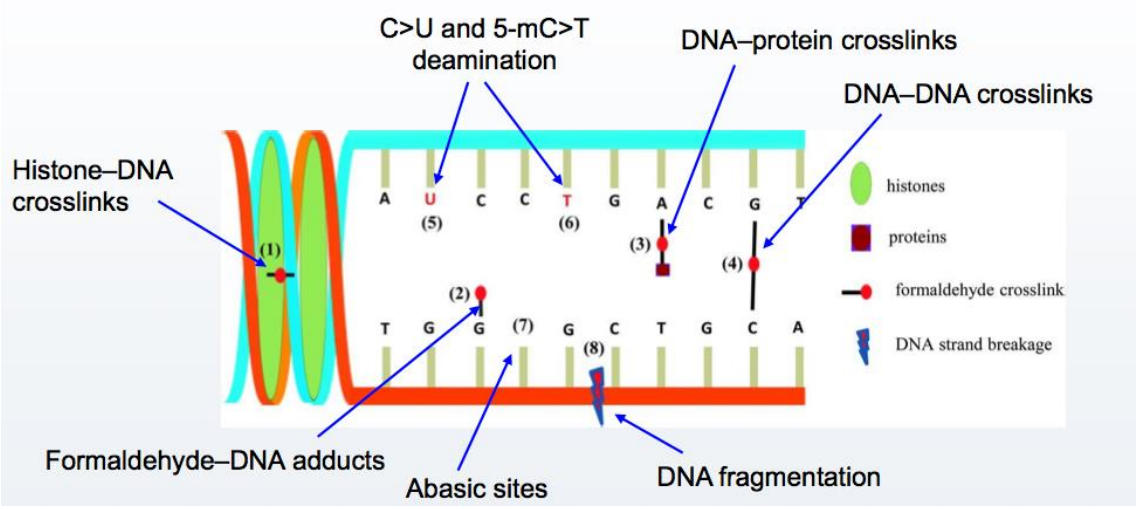


Figure 23 - Known effects of fixation process in DNA quality
Formaldehyde, the principal constituent of formalin, is highly reactive with DNA bases and proteins leading to several types of damages which are found in DNA extracted from formalin-fixed tissues. As such, the DNA loss of quality can be due to histone–DNA crosslinks (1); formaldehyde–DNA adducts (2); DNA–protein crosslinks (3); DNA–DNA crosslinks (4); deamination of cytosine and 5-mC resulting in uracil (5) and thymine (6), respectively; loss of DNA bases resulting in abasic sites (7); fragmentation of DNA (8) due to DNA strand breakage. (figure adapted from (92))

Concerning the disparities in the number of variants detected in the different genes, the results obtained were quite the expected: bigger genes had generally more variants pointed by NGS technology. Also, these same genes (and not only) showed multiple damaging mutation sites simultaneously. However, some of these findings were analysed and all came out to be false positives. As such, this information of multiple variants close to each other as detected by the IGV can be used as an indicator of false positive data (e.g. fixation artefact C>T and G>A transitions). Nonetheless, further studies are needed to continue the evaluation of these simultaneous mutations because if any turns out to really be positive, it would be a bright new challenge to deeper understand glioma's physiopathogenesis, because a question stands: why should a gene have more than one damaging mutation?

Concerning costs, expenses for the 19-gene panel NGS was comparable to all applied PCR/Sanger sequencing/pyrosequencing assays.

The central limitation of this study, besides the short period of research time, relates to the fact that the present gene panel only assesses 19 glioma-related genes, not covering certain genetic abnormalities known to also contribute to the physiopathogenesis of some glioma types, such as structural alterations like 1p19q codeletion in oligodendrogliomas (14,90), gene amplifications (e.g. *RB1*, *EGFR*, *CDK4*, *CDK6*, *PDGFRA* in glioblastomas

(87)) and gene fusions (e.g. *KIAA1549-BRAF* in pilocytic astrocytomas (93); RELA in ependymomas (14,94)). However, these assessments would be easily achieved by: (i) complementing the gene panel with additional glioma candidate genes when diagnostically required allowing mutational and copy number analysis (42) and (ii) in case of gene fusions detection, specifically designing panels based on transcriptomic sequencing (63).

The same way diagnostic and prognostic biomarkers may be detected with NGS gene panel approach, individualized therapy strategies can be feasible through molecular analysis resorting to NGS gene panels covering the potential targets of the available drugs. Currently, the only clinically relevant predictive biomarker for malignant gliomas which is capable of guiding therapeutic stratification is O6-methylguanine DNA methyltransferase (*MGMT*) promoter methylation in GBM patients – *MGMT* + leads to an increased response to Temozolomide treatment (11). However, the assessment of *MGMT* methylation status requires independent analysis, which is also a limitation of this NGS-based approach.

Hence, despite NGS promising applicability in detection of diagnostic, prognostic and targeted treatment biomarkers and its great results in the present study, further studies to standardize pipelines for filtering and for interpretation of genetic variations critical for clinical decision-making are still crucial to enhance NGS safe use in routine diagnostics of brain tumours.

6 CONCLUSIONS

WHO 2016 classification of CNS tumours have recently included molecular parameters in addition to histological features for establishing the diagnostics of gliomas, enhancing the increased importance of these in the clinical outcomes and, so, in the tumour management. Currently, single-gene approach technologies are the gold standard to perform the molecular diagnostics of gliomas. However, NGS gene panels-approach provides the opportunity to analyse a large set of molecular targets in a massively paralleled way with a low DNA material input and, as such, may become a valuable tool in future glioma diagnostics providing a cost and time efficient alternative to currently used conventional pyro- and Sanger sequencing techniques.

Hence, the aim of this study was to apply and verify if NGS could effectively replace the conventional single-gene approach techniques to perform the molecular diagnostics of gliomas. Therefore, it was established and verified a customized NGS gene-panel for diagnostic sequencing of 19 genes frequently aberrant in gliomas allowing to assess NGS ability to detect well-known hotspot mutations and to correctly identify new-mutations through a comparison of its results with the gold-standard sequencing methods for molecular diagnostics of gliomas.

The process of DNA preparation and sequencing worked marvellously, carrying a very good quality score. It took only one week of preparation of 30 libraries and three days of sequencing them and obtaining the outputs to collect hundreds of genetic variants for each sample. The filtering process had a huge impact in data analysis by reducing in a great amount the NGS outputs to further evaluation. Library preparation also showed to considerably influence the DNA sequence changes detected by NGS in which samples with better libraries generally presented higher percentage of reads identified and less variants detected. Moreover, the filtering process exhibited different impact in the results of different quality libraries, confining more effectively the data of the better ones. NGS-gene panel showed 100% concordance with the pyrosequencing results previously and further performed concerning the detection of diagnostically relevant DNA abnormalities such as hotspot mutations in *IDH1*, *H3F3A*, *HIST1H3B*, *BRAF* and *TERT* promotor. On the other hand, the validation of new-mutations detected by NGS displayed a poorer pattern, only confirming 34% by Sanger sequencing analysis. Nonetheless, the majority of the variants assessed were SNVs C:G > T:A and, as such, FFPE artefacts were certainly

the main source of most false positives. Further, genes with higher size showed generally a higher number of variants detected by NGS and they commonly presented multiple damaging mutation sites simultaneously which all came out to be false positives in Sanger sequencing analysis leading to conclude that this information can be further used as an indicator of false positive data (e.g. fixation artefact).

Despite the large amount of filtered data, hundreds remained to be analysed. More time would be needed to continue the verification of new-mutations detected by NGS allowing stronger and more reliable conclusions. Anyway, further studies are crucial to continue to try to find the best way to filter NGS outputs in order to collect mostly the real and clinically relevant variants, which is still the major obstacle of NGS use in routine molecular diagnostics. Additionally, and trying to overcome NGS difficulty to identify FFPE-DNA artefacts, a treatment of the FFPE-DNA with uracil-glycosylase and a better pattern of mutation selection to further validation can be used in future work to more accurately assess the ability of NGS to detect real new-mutations.

Due to the fact that the NGS-gene panel studied only assessed 19 glioma-related genes and couldn't evaluate certain genetic abnormalities important in glioma pathophysiology, future research complementing the gene panel and allowing copy number analysis and gene fusion detection would enrich the present conclusions concerning NGS ability to correctly detect genetic abnormalities valuable for glioma molecular diagnostics and would improve even more the knowledge of the molecular pathology of gliomas.

In summary, with the majority of good results obtained, one can point out that NGS-based approach is indeed a potential substitute for the conventional technologies to perform the molecular diagnostics of brain tumours. However, despite NGS immense ability to furnish important sequencing data in a time and cost more efficient way than conventional technologies, it is not possible to argue its application in glioma routine molecular diagnostics at the current time due to its cloudy results concerning the new-mutations detected. Consequently, NGS safe use in its wide range of potential clinical applications (diagnostics, prognostics, targeted treatment) must be delayed until it is established clear standard pipelines for filtering and for interpretation of genetic variations critical for clinical decision-making. Hence, it is crucial to pursue the path of extended research of this NGS technology in order to increase the quality and accuracy of its use and, then, to be possible to enjoy all its advantages respecting to the classical methods.

References

1. Ho VKY, Reijneveld JC, Enting RH, Bienfait HP, Robe P, Baumert BG, et al. Changing incidence and improved survival of gliomas. Eur J Cancer. 2014;50(13):2309–18.
2. IARC. Cancer of the brain and central nervous system [Internet]. [cited 2017 Oct 6]. Available from: http://eco.iarc.fr/eucan/Cancer.aspx?Cancer=34
3. Visser O, Ardanaz E, Botta L, Sant M, Tavilla A, Minicozzi P. Survival of adults with primary malignant brain tumours in Europe; Results of the EUROCARE-5 study. Eur J Cancer. 2015;51(15):2231–41.
4. Ekman M, Westphal M. Cost of brain tumour in Europe. Eur J Neurol. 2005;12(January):63–7.
5. De Robles P, Fiest KM, Frolkis AD, Pringsheim T, Atta C, St. Germaine-Smith C, et al. The worldwide incidence and prevalence of primary brain tumors: A systematic review and meta-analysis. Neuro Oncol. 2015;17(6):776–83.
6. Mao H, LeBrun DG, Yang J, Zhu VF, Li M. Deregulated Signaling Pathways in Glioblastoma Multiforme: Molecular Mechanisms and Therapeutic Targets. Cancer Invest. 2012;30(1):48–56.
7. Bastien JIL, McNeill KA, Fine HA. Molecular characterizations of glioblastoma, targeted therapy, and clinical results to date. Cancer. 2015;121(4):502–16.
8. Rasmussen BK, Hansen S, Laursen RJ, Kosteljanetz M, Schultz H, Nørgård BM, et al. Epidemiology of glioma: clinical characteristics, symptoms, and predictors of glioma patients grade I–IV in the the Danish Neuro-Oncology Registry. J Neurooncol. 2017;1–9.
9. Immanuel V, Kingsley PA, Negi P, Isaacs R, Grewal SS. Variegated colors of pediatric glioblastoma multiforme: what to expect? Rare Tumors. 2017;9(2).
10. Hanif F, Muzaffar K, Perveen K, Malhi SM, Simjee SU. Glioblastoma Multiforme: A Review of its Epidemiology and Pathogenesis through Clinical Presentation and Treatment. Asian Pac J Cancer Prev. 2017;18(1):3–9.

11. Weller M, van den Bent M, Tonn JC, Stupp R, Preusser M, Cohen-Jonathan-Moyal E, et al. European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. Lancet Oncol. 2017;18(6):e315–29.

12. Arbor A, Aires B. Recent advances and future of immunotherapy for glioblastoma. 2016;16(10):1245–64.

13. Wesseling P, van den Bent M, Perry A. Glioblastoma: pathology, molecular mechanisms and markers. Acta Neuropathol. 2015;129(6):809–27.

14. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 2016;131(6):803–20.

15. Miller JJ, Shih HA, Andronesi OC, Cahill DP. Isocitrate dehydrogenase-mutant glioma: Evolving clinical and therapeutic implications. Cancer. 2017;1–12.

16. Chen R, Smith-Cohn M, Cohen AL, Colman H. Glioma Subclassifications and Their Clinical Significance. Neurotherapeutics. 2017;14(2):284–97.

17. Karsy M, Guan J, Cohen AL, Jensen RL, Colman H. New Molecular Considerations for Glioma: IDH, ATRX, BRAF, TERT, H3 K27M. Curr Neurol Neurosci Rep. 2017;17(2).

18. Masui K, Mischel PS, Reifenberger G. Molecular classification of gliomas. 1st ed. Vol. 134, Handbook of Clinical Neurology. Elsevier B.V.; 2016. 97-120 p.

19. Kasner E, Hunter CA, Ph D, Kariko K, Ph D. Epidemiologic and Molecular Prognostic Review of Glioblastoma. 2013;70(4):646–56.

20. Meyer-Puttlitz B, Hayashi Y, Waha A, Rollbrocker B, Boström J, Wiestler OD, et al. Molecular genetic analysis of giant cell glioblastomas. Am J Pathol. 1997;151(3):853–7.

21. Soomro SH, Ting LR, Qing YY, Ren M. Molecular biology of glioblastoma: Classification and mutational locations. J Pak Med Assoc. 2017;67(9):1410–4.

22. Schwartzbaum JA, Fisher JL, Aldape KD, Wrensch M. Epidemiology and molecular pathology of glioma. Nat Clin Pract Neurol. 2006;2(9):494–503.

23. McNeill KA. Epidemiology of Brain Tumors. Neurol Clin. 2016;34(4):981–98.

24. Brenner A V., Linet MS, Fine HA, Shapiro WR, Selker RG, Black PM, et al. History of allergies and autoimmune diseases and risk of brain tumors in adults. Int J Cancer. 2002;99(2):252–9.

25. Joice R, Nilsson SK, Montgomery J, Dankwa S, Morahan B, Seydel KB, et al. Effects of antihistamine and anti-inflammatory medication use on risk of specific glioma histologies. 2014;6(244):1–16.

26. Benson VS, Pirie K, Schüz J, Reeves GK, Beral V, Green J. Mobile phone use and risk of brain neoplasms and other cancers: Prospective study. J Immunol. 2015;195(5):792–802.

27. Deltour I, Auvinen A, Feychting M, Johansen C, Klaeboe L, Sankila R, et al. Mobile Phone Use and Incidence of Glioma in the Nordic Countries 1979–2008. Epidemiology. 2012;23(2):301–7.

28. Oliveira AI, Anjo SI, Vieira de Castro J, Serra SC, Salgado AJ, Manadas B, et al. Crosstalk between glial and glioblastoma cells triggers the "go-or-grow" phenotype of tumor cells. Cell Commun Signal. 2017;15(1):37.

29. Stupp R, Mason WP, Van Den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. N Engl J Med. 2005;35210:987–96.

30. Oldridge DA, Wood AC, Weichert-leahey N, Crimmins I, Winter C, Mcdaniel LD, et al. Emerging immunotherapies for glioblastoma. 2016;528(7582):418–21.

31. Takahashi Y, Makino K, Nakamura H, Hide T, Yano S, Kamada H, et al. Clinical characteristics and pathogenesis of cerebellar glioblastoma. Mol Med Rep. 2014;10(5):2383–8.

32. Faguer R, Tanguy JY, Rousseau A, Clavreul A, Menei P. Early presentation of primary glioblastoma. Neurochirurgie. 2014;60(4):188–93.

33. Nelson S, Cha S. Imaging glioblastoma multiforme. Cancer J. 2003;9(2):134–45.

34. Agnihotri S, Burrell KE, Wolf A, Jalali S, Hawkins C, Rutka JT, et al. Glioblastoma, a brief review of history, molecular genetics, animal models and novel therapeutic strategies. Arch Immunol Ther Exp (Warsz). 2013;61(1):25–41.

35. Itakura H, Achrol AS, Mitchell LA, Loya JJ, Liu T, Erick M, et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. 2016;7(303).

36. Rong Y, Durden DL, Van Meir EG, Brat DJ. "Pseudopalisading" Necrosis in Glioblastoma: A Familiar Morphologic Feature That Links Vascular Pathology, Hypoxia, and Angiogenesis. J Neuropathol Exp Neurol. 2006;65:529–539.

37. Brat DJ, Castellano-Sanchez AA, Hunter SB, Pecot M, Cohen C, Hammond EH, et al. Pseudopalisades in Glioblastoma Are Hypoxic, Express Extracellular Matrix Proteases, and Are Formed by an Actively Migrating Cell Population. Cancer Res. 2004;64(3):920–7.

38. Nakada M, Kita D, Watanabe T, Hayashi Y, Teng L, Pyko I V., et al. Aberrant signaling pathways in Glioma. Cancers (Basel). 2011;3(3):3242–78.

39. Wilson T, Karajannis M, Harter D. Glioblastoma multiforme: State of the art and future therapeutics. Surg Neurol Int. 2014;5(1):64.

40. Blumenthal DT, Dvir A, Lossos A, Tzuk-Shina T, Lior T, Limon D, et al. Clinical utility and treatment outcome of comprehensive genomic profiling in high grade glioma patients. J Neurooncol. 2016;130(1):211–9.

41. De Leng WWJ, Gadellaa-Van Hooijdonk CG, Barendregt-Smouter FAS, Koudijs MJ, Nijman I, Hinrichs JWJ, et al. Targeted next generation sequencing as a reliable diagnostic assay for the detection of somatic mutations in tumours using minimal DNA amounts from formalin fixed paraffin embedded material. PLoS One. 2016;11(2):1–18.

42. Felsberg J, Malzkorn B, Reifenberger G. Molecular Diagnostics of Gliomas Using Next Generation Sequencing of a Glioma-Tailored Gene Panel. 2017;27:146–59.

43. Buczkowicz P, Hawkins C. Pathology, Molecular Genetics, and Epigenetics of Diffuse Intrinsic Pontine Glioma. Front Oncol. 2015;5(June):1–9.

44. Lapin DH, Tsoli M, Ziegler DS. Genomic Insights into Diffuse Intrinsic Pontine Glioma. Front Oncol. 2017;7(March).

45. Chamdine O, Gajjar A. Molecular characteristics of pediatric high-grade gliomas. CNS Oncol. 2014;3(6):433–43.

46. Gelaye B, Rondon M, Araya PR, A PM. Unique genetic and epigenetic mechanisms driving signatures of paediatric diffuse high-grade glioma. 2016;3(10):973–82.

47. Bush NAO, Butowski N. The Effect of Molecular Diagnostics on the Treatment of Glioma. Curr Oncol Rep. 2017;19(4).

48. Jiao Y, Killela PJ, Reitman ZJ, Rasheed AB, Heaphy CM, de Wilde RF, et al. Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas. Oncotarget. 2012;3(7):709–22.

49. Takahashi Y, Akahane T, Sawada T, Ikeda H, Tempaku A, Yamauchi S, et al. Adult classical glioblastoma with a BRAF V600E mutation. World J Surg Oncol. 2015;13:100.

50. Gleize V, Alentorn A, Connen De K??rillis L, Labussi??re M, Nadaradjane AA, Mundwiller E, et al. CIC inactivating mutations identify aggressive subset of 1p19q codeleted gliomas. Ann Neurol. 2015;78(3):355–74.

51. Andersson U, Schwartzbaum J, Wiklund F, Sjostrom S, Liu Y, Tsavachidis S, et al. A comprehensive study of the association between the EGFR and ERBB2 genes and glioma risk. Acta Oncol. 2010;49(6):767–75.

52. Xu GF, Xie WF. Effect of ERBB2 expression on invasiveness of glioma TJ905 cells. Asian Pac J Trop Med. 2013;6(12):964–7.

53. Lehtinen B, Raita A, Kesseli J, Annala M, Nordfors K, Yli-Harja O, et al. Clinical association analysis of ependymomas and pilocytic astrocytomas reveals elevated FGFR3 and FGFR1 expression in aggressive ependymomas. BMC Cancer. 2017;17(1):310.

54. Gouazé-Andersson V, Delmas C, Taurand M, Martinez-Gala J, Evrard S, Mazoyer S, et al. FGFR1 induces glioblastoma radioresistance through the PLCg/Hif1a pathway. Cancer Res. 2016;76(10):3036–44.

55. Yang L, Zhu J ya, Zhang J guo, Bao B jun, Guan C qi, Yang X jing, et al. Far upstream element-binding protein 1 (FUBP1) is a potential c-Myc regulator in esophageal squamous cell carcinoma (ESCC) and its expression promotes ESCC progression. Tumor Biol. 2016;37(3):4115–26.

56. Castel D, Philippe C, Calmon R, Le Dret L, Truffaux N, Boddaert N, et al. Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes. Acta Neuropathol. 2015;130(6):815–27.

57. Cloughesy TF, Cavenee WK, Mischel PS. Glioblastoma: From Molecular Pathology to Targeted Treatment. Annu Rev Pathol Mech Dis. 2014;9(1):1–25.

58. Mato Prado M, Frampton AE, Stebbing J, Krell J. Gene of the month: PIK3CA. J Clin Pathol. 2015;68(10):763–5.

59. Bleeker FE, Lamba S, Zanon C, Molenaar RJ, Hulsebos TJ, Troost D, et al. Mutational profiling of kinases in glioblastoma. BMC Cancer. 2014;14(1):718.

60. Wang P, Ye JA, Hou CX, Zhou D, Zhan SQ. Combination of lentivirus-mediated silencing of PPM1D and temozolomide chemotherapy eradicates malignant glioma through cell apoptosis and cell cycle arrest. Oncol Rep. 2016;36(5):2544–52.

61. Zhang L, Chen LH, Wan H, Yang R, Wang Z, Feng J, et al. Exome sequencing identifies somatic gain-of-function PPM1D mutations in brainstem gliomas. 2015;46(7):726–30.

62. McNamara M, Sahebjam S, Mason W. Emerging Biomarkers in Glioblastoma. Cancers (Basel). 2013;5(3):1103–19.

63. Nikiforova MN, Wald AI, Melan MA, Roy S, Zhong S, Hamilton RL, et al. Targeted next-generation sequencing panel (GlioSeq) provides comprehensive genetic profiling of central nervous system tumors. Neuro Oncol. 2016;18(3):379–87.

64. Sahm F, Schrimpf D, Jones DTW, Meyer J, Kratz A, Reuss D, et al. Next-generation sequencing in routine brain tumor diagnostics enables an integrated diagnosis and identifies actionable targets. Acta Neuropathol. 2016;131(6):903–10.

65. Garrido-Cardenas JA, Garcia-Maroto F, Alvarez-Bermejo JA, Manzano-Agugliaro F. DNA sequencing sensors: An overview. Sensors (Switzerland). 2017;17(3):1–15.

66. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016;107(1):1–8.

67. Kamps R, Brandão R, Bosch B, Paulussen A, Xanthoulea S, Blok M, et al. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. Int J Mol Sci. 2017;18(2):308.

68. Meldrum C, Doyle M a, Tothill RW. Next-Generation Sequencing for Cancer Diagnostics: a Practical Perspective. Clin Biochem Rev. 2011;32(4):177–95.

69. Yang Y, Xie B, Yan J. Application of Next-generation Sequencing Technology in Forensic Science. Genomics Proteomics Bioinformatics. 2014;12(5):190–7.

70. Technologies HS. High-Throughput Sequencing Technologies. 2016;58(4):586–97.

71. Van Dijk EL, Lè Ne Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30:418–26.

72. Illumina. An Introduction to Next-Generation Sequencing Technology. Illumina.com. 2015;(illumina):1–16.

73. Dumur CI, Almenara JA, Powers CN, Ferreira-Gonzalez A. Quality control material for the detection of somatic mutations in fixed clinical specimens by next-generation sequencing. Diagn Pathol. 2015;10(1):169.

74. Sheerin U. The the Use of Next of Generation Parkinson ' s Sequencing Technologies to Dissect A etiologies disease and Dystonia. 2014;(March):1–225.

75. Illumina. Calculating percent passing filter for patterned and non-patterned flow cells: A comparison of methods for calculating percent passing filter on Illumina flow cells. 2015;

76. Qiagen. Pyrosequencing Technology and Platform Overview -. Qiagen. 2013;5–7.

77. Delaney C, Arbor A. Immunosenecence. 2015;1343:249–64.

78. Appell ML. Pharmacogenetic studies of thiopurines: focus on thiopurine methyltransferase. 2005.

79. ThermoFisherScientific. Sanger sequencing technology overview [Internet]. [cited 2017 Jul 6]. Available from: https://www.thermofisher.com/pt/en/home/life-science/sequencing/sanger-sequencing/sanger_sequencing_method.html

80. Li Y, Zhou X, editors. 2.5.2 Sanger Sequencing. In: Atlas of Oral Microbiology. 2015. p. 15–40.

81. Saumya S. Jamuar, Alissa M. D'Gama CAW, editor. Section I: The Genome Tools and Methods. In: Genomics, Circuits, and pathways in clinical neuropsychiatry. 2016. p. 179–99.

82. Sanger F, Nicklen S. DNA sequencing with chain-terminating. 1977;74(12):5463–7.

83. GATC-Biotec. Sanger sequencing [Internet]. Sanger sequencing. 2017 [cited 2017 Jul 8]. Available from: detection of labelled chain-terminating nucleotides that are incorporated by a DNA polymerase during the replication of a template.

84. Bechet D, Gielen GGH, Korshunov A, Pfister SM, Rousso C, Faury D, et al. Specific detection of methionine 27 mutation in histone 3 variants (H3K27M) in fixed tissue from high-grade astrocytomas. Acta Neuropathol. 2014;128(5):733–41.

85. Capper D, Preusser M, Habel A, Sahm F, Ackermann U, Schindler G, et al. Assessment of BRAF V600E mutation status by immunohistochemistry with a mutation-specific monoclonal antibody. Acta Neuropathol. 2011;122(1):11–9.

86. Capper D, Zentgraf H, Balss J, Hartmann C, Von Deimling A. Monoclonal antibody specific for IDH1 R132H mutation. Acta Neuropathol. 2009;118(5):599–601.

87. Riemenschneider MJ, Jeuken JWM, Wesseling P, Reifenberger G. Molecular diagnostics of gliomas: State of the art. Acta Neuropathol. 2010;120(5):567–84.

88. Devereux HL. Automated DNA sequencing. Methods Mol Med. 1999;31:55–61.

89. Sahm F, Koelsche C, Meyer J, Pusch S, Lindenberg K, Mueller W, et al. CIC and FUBP1 mutations in oligodendrogliomas, oligoastrocytomas and astrocytomas. Acta Neuropathol. 2012;123(6):853–60.

90. Yip S, Butterfield YS, Morozova O, Chittaranjan S, Michael D, An J, et al. Concurrent CIC mutations, IDH mutations and 1p/19q loss distinguish oligodendrogliomas from other cancers. 2013;226(1):7–16.

91. Do H, Dobrovic A. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil- DNA glycosylase. Oncotarget. 2012;3(5):546–58.

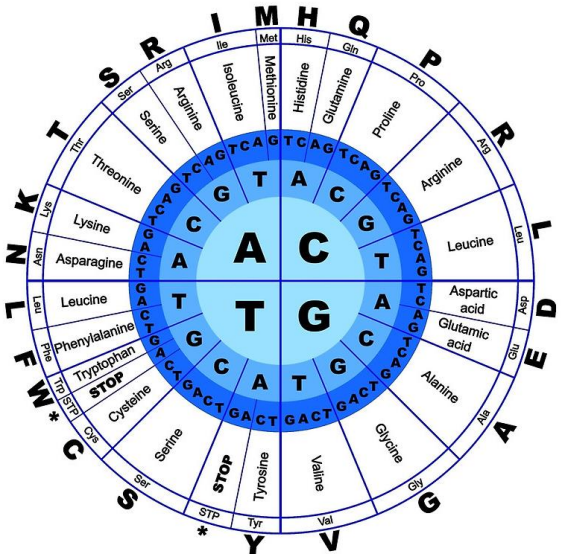
92. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: Causes and strategies for minimization. Clin Chem. 2015;61(1):64–71.

93. Faulkner C, Ellis HP, Shaw A, Penman C, Palmer A, Wragg C, et al. BRAF Fusion Analysis in Pilocytic Astrocytomas: KIAA1549-BRAF 15-9 Fusions Are More Frequent in the Midline Than Within the Cerebellum. J Neuropathol Exp Neurol. 2015;74(9):867–72.

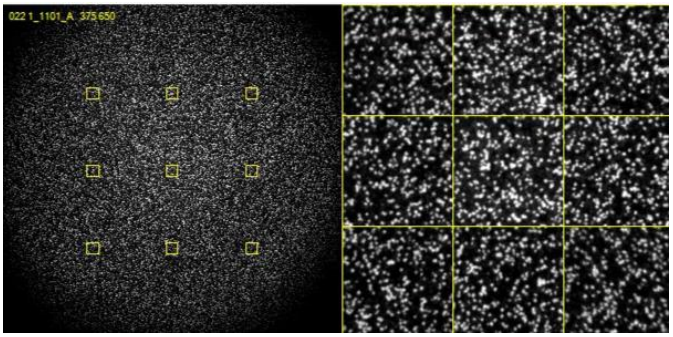
94. Parker M, Mohankumar KM, Punchihewa C, Weinlich R, Dalton JD, Li Y, et al. C11orf95-RELA fusions drive oncogenic NF-κB signaling in ependymoma. 2014;506(7489):451–5.

Appendixes

A.1 Genetic code

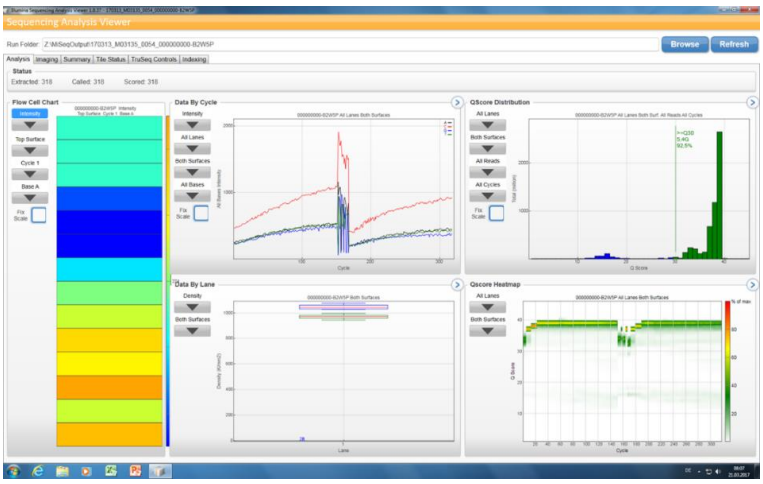


A.2 Flow cell image during sequencing process



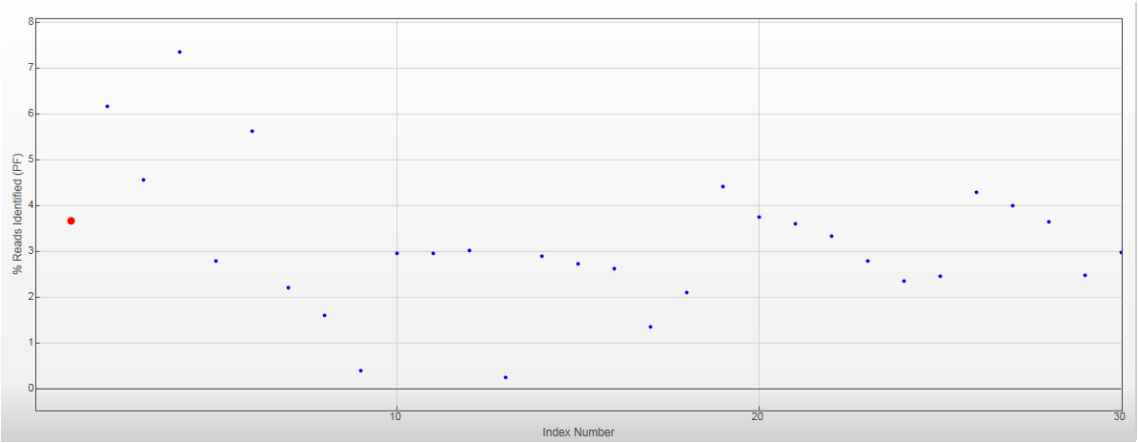
A.3 Illumina Sequencing Analysis Viewer (SAV): software landscape

Application that shows the critical quality metrics generated by the MiSeq real-time analysis software.



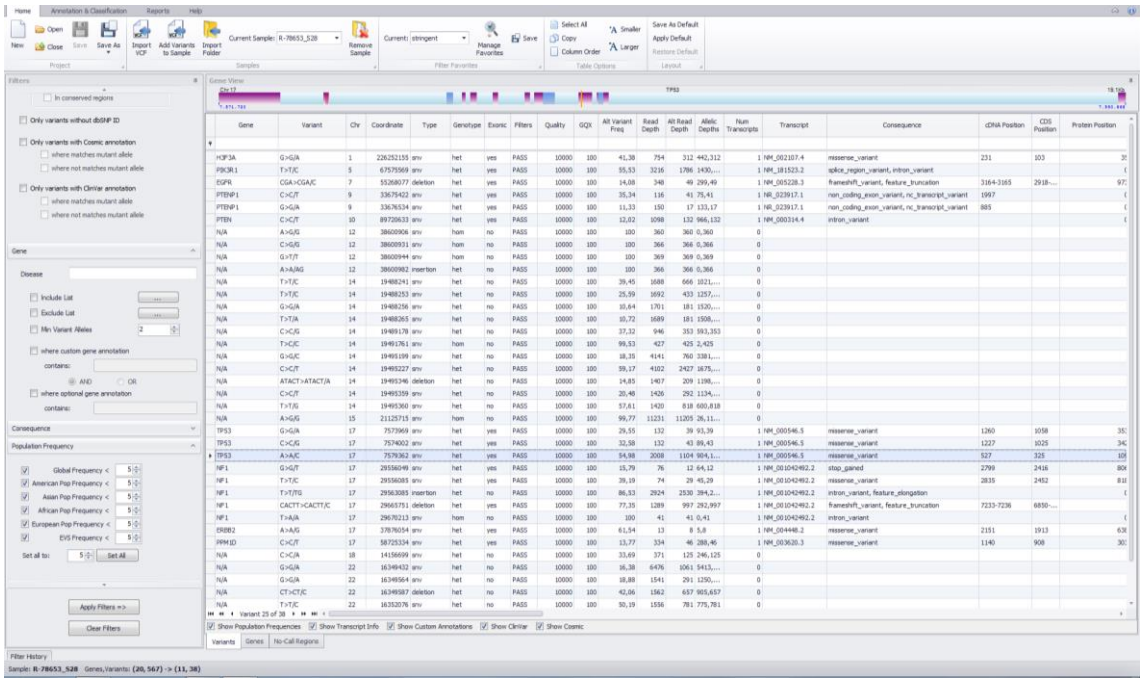
A.4 Reads mapped to indexes from Illumina Sequencing Analysis Viewer

Percentage of identified reads for each index number/sample.



A.5 Illustration of Illumina VariantStudio Variant Analysis Software

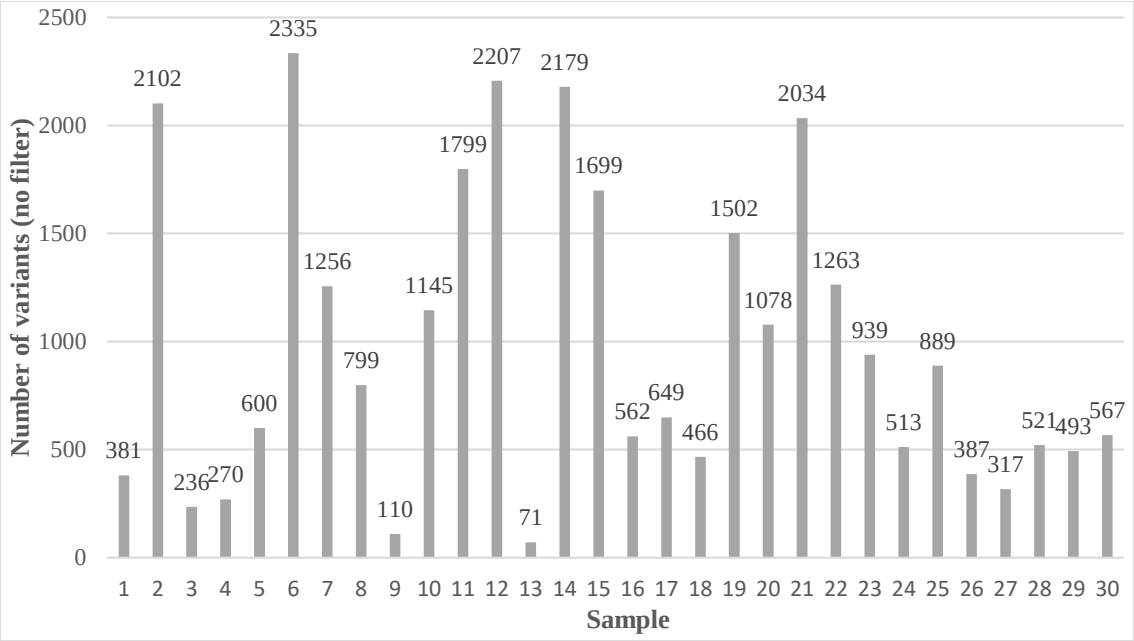
This software allows us to choose the sample to further analyse and set diverse filtering parameters. It displays, among others, an illustrative representation of the genetic variant within the gene (exons, introns and site of mutation clearly designed), the gene abbreviation, the variant type, the genomic location, the variant frequency, the read depths, the consequence and if it is deleterious/damaging or tolerated/benign. Also, this software is connected with several databases (e.g. COSMIC) allowing a quicker perception if the mutation is or not already described.



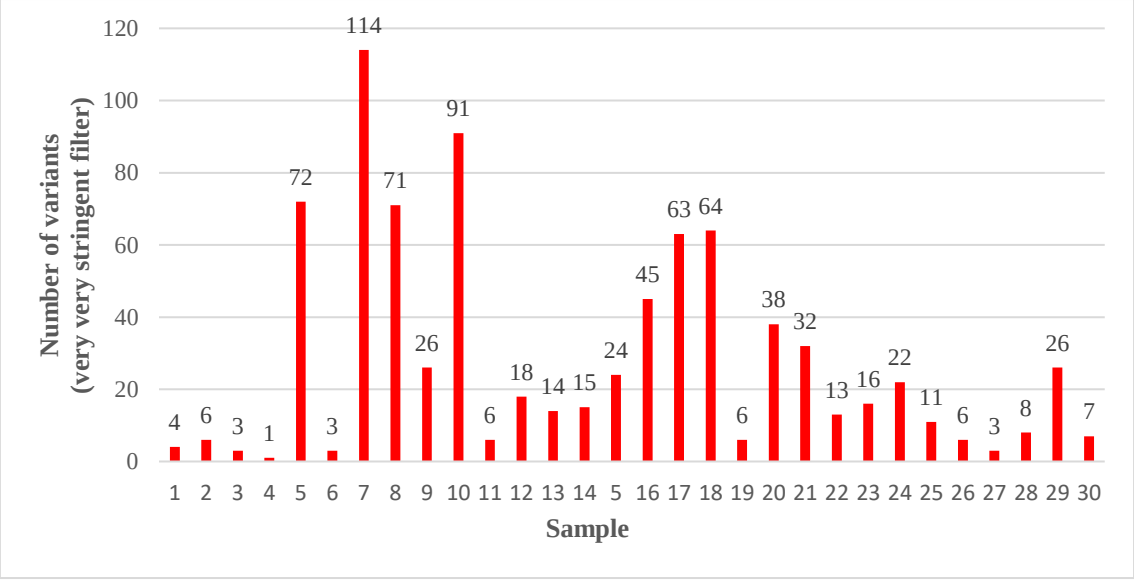
A.6 Number of variants per sample detected by NGS technology

NGS results showing the difference in the number of variants obtained for each sample when no filter **A)** *vs* the most stringent filter **B)** were applied.

A) No filter



B) Very very stringent filter



A.7 Primers designed for validation of the selected NGS novel mutations with Sanger sequencing

Several of the primers listed, which were designed for further confirmation of the new-mutations detected by NGS, were capable of assessing more than one mutation detected.

a. Primers designed for amplification of gDNA for the first set of Sanger sequencing validations of the selected new mutations detected by NGS

Sample ID	*Gene amplified*	*Primer sequence*
4099	***HIST1H3B***	fw: 5' TCTTGCGAGCAGCCTTGG 3' rev: 5' AGACGTCTCTGCAGGCAAGC 3'
4099	***IDH1***	fw: 5' TTACCCATCCACTCACAA 3' rev: 5' GAAGCATAATGTTGGCGTC 3'
4101	***H3F3A***	fw: 5' TCAATGCTGGTAGGTAAG 3' rev: 5' TTGTAGCCAGTTGCTTCC 3'
4101	***FUBP1***	fw: 5' AGTACTCTTCCCAAGCCTTG 3' rev: 5' ACCCTCACTGTCACATTGCA 3'
4101	***FUBP1***	fw: 5' TCCCAGTTGGCACAATAA 3' rev: 5' TAATCCTGGTGGACCTGGAC 3'
4101	***FUBP1***	fw: 5' TGCGAACACCAGCATCAT 3' rev: 5' AACCCTGAGAAGCTAGCTT 3'
4101	***FUBP1***	fw: 5' TACCATGAGAATGTAACA 3' rev: 5' CTAACTAGTAATTGGCAGAGG 3'
4101	***HIST1H3B***	fw: 5' TTGGTAGCCAGCTGCTTGCGT 3' rev: 5' ACGTCTCTGCAGGCAAGCTT 3'
4101	***EGFR***	fw: 5' TCTGTCACTGACTGCTGTGA 3' rev: 5' ATTCCCTGCCTCGGCTGACATT 3'
4101	***BRAF***	fw: 5' GTAACTCAGCAGCATCTCA 3' rev: 5' TTGGTCTAGCTACAGTGA 3'
4101	***PPM1D***	fw: 5' GATACAGATGTAGTGGCAGC 3' rev: 5' ATCTGCTCGGAGCATACGCT 3'
4101	***CIC***	fw: 5' ACAGCTCACCCTGGCCTAT 3' rev: 5' ACCTTCTGAGGCTGAGAGGT 3'
4101	***CIC***	fw: 5' AGCCAGGCTGGAACAGTCA 3' rev: 5' CCCTCACCTGAGAGCAGG 3'
4101	***CIC***	fw: 5' AGCTGCCGCCTGCCTGTG 3' rev: 5' AGAATGGTGGCTGCAGGT 3'

4102	***IDH1***	fw: 5’ GCCAACATGACTTACTTGATCC 3’ rev: 5’ AATATTCTGGGTGGCACGGTC 3’
4107	***PIK3R1***	fw: 5’ TCTAGGATCAAGTTGTCA 3’ rev: 5’ AACTCACCTGGGATGTGC 3’
4107	***PIK3R1***	fw: 5’ GCCAATATTCACTGGTGGAAG 3’ rev: 5’ AAGTGCCATCTCGCTTCCCT 3’
R-78610	***CIC***	fw: 5’ AGCCACTGCCACTGGTGA 3’ rev: 5’ ACATCCAGCAGGTAGAGAG 3’

b. Primers designed for amplification of gDNA for the second set of Sanger sequencing validations of the selected new mutations detected by NGS

Sample ID	*Gene amplified*	*Primer sequence*
585	***PTEN***	fw: 5’ TTGCAAATGTTTAACATAGG 3’ rev: 5’ ACTGACCTTAAAATTTGGAG 3’
588	***TP53***	fw: 5’ AATCAGTGAGGAATCAGAGG 3’ rev: 5’ CAAGATGTTTTGCCAACTGG 3’
588	***NF1***	fw: 5’ CATAAAATTACCCAAGTTGC 3’ rev: 5’ TGTTATAGTTGGGCAAGAGG 3’
588	***ATRX***	fw: 5’ TCTTCCTCTTCCTTAAGAGC 3’ rev: 5’ ATGACTTTGTACTGTTTACC 3’
4118	***TP53***	fw: 5’ AAGGGACAGAAGATGACAGG 3’ rev: 5’ GGATGATTTGATGCTGTCC 3’
4118	***NF1***	fw: 5’ TTGTCCACATTAGGCTTAGG 3’ rev: 5’ AGAACAAGGAACCACATTGG 3’
4119	***TP53***	fw: 5’ TTGCTTACCTCGCTTAGTGC 3’ rev: 5’ ATTTCCTTACTGCCTCTTGC 3’
4119	***ATRX***	fw: 5’ ACCCTTTCTTCTGTTTCTGC 3’ rev: 5’ TCTGAAGAATCTAAGAAGCC 3’
4120	***ERBB2***	fw: 5’ TCGCTCACAACCAAGTGAGG 3’ rev: 5’ GGTTTCAATGACGGTGAAGG 3’
4120	***TP53***	fw: 5’ CTCCTGACCTGGAGTCTTCC 3’ rev: 5’ CTTGGGCCTGTGTTATCTCC 3’
4121	***NF1***	fw: 5’ GTGGGTCTAGAATTGAGTCC 3’ rev: 5’ TCCTAAGAGGCAAGCTGACC 3’
4121	***NF1***	fw: 5’ CTGGTTATATCTGCATTAGG 3’ rev: 5’ GGAGAAGCTGAAATAGAACC 3’
4121	***NF1***	fw: 5’ TTATTTCTGGACAGTCTACG 3’ rev: 5’ GCAAGTCCTATGAACTTATC 3’
4121	***NF1***	fw: 5’ TTGCTTACGACAACGTCTCC 3’ rev: 5’ TGTGGAATACCTTCAGGTCC 3’

4121	***TP53***	fw: 5’ ACTGACAACCACCCTTAACC 3’ rev: 5’ TCAGCATCTTATCCGAGTGG 3’
4121	***CIC***	fw: 5’ CTGAGTCTGCTTCTGTTTGG 3’ rev: 5’ TCACTCTCTAACCGCCTTCC 3’
4121	***ATRX***	fw: 5’ GGGAACCCTCAACAACAAGG 3’ rev: 5’ CGTTTCAACATACCAACTGG 3’
4121	***ATRX***	fw: 5’ CAACTGAACTCTGAACTTCC 3’ rev: 5’ TCAGAATGTTCCAACAGAGG 3’
4125	***NF1***	fw: 5’ CTAGCAGAAATTATATCAATGAG 3’ rev: 5’ AGACAAGCTATGTCTTGACC 3’
4125	***ACVR1***	fw: 5’ ACTGTCCATTCTTCTTAACC 3’ rev: 5’ CATTATCATGAAATGGGATCG 3’
4125	***ATRX***	fw: 5’ TCAGGTAACTTTTCAGTGCC 3’ rev: 5’ AGAGCAAGCATCTCAAAACC 3’
4125	***H3F3A***	fw: 5’ AAATCGACCGGTGGTAAAGC 3’ rev: 5’ TACAAGAGAGACTTTGTCCC 3’
4125	***PTEN***	fw: 5’ TCCTTTTGAAGACCATAACC 3’ rev: 5’ AAACCCAAAATCTGTTTTCC 3’
4125	***TP53***	fw: 5’ GCAAATGCCCCAATTGCAGG 3’ rev: 5’ AAGCGAGGTAAGCAAGCAGG 3’
4129	***TP53***	fw: 5’ CTGGAGGGCCACTGACAACC 3’ rev: 5’ GTCCCCAGGCCTCTGATTCC 3’
4129	***EGFR***	fw: 5’ TCTTCCAGTGTTCTAATTGC 3’ rev: 5’ AACACAGTGACATGAGATGC 3’
4129	***NF1***	fw: 5’ AAGTCGTCATGTCACTTAGG 3’ rev: 5’ AACCCACTAATACTTGAAGG 3’
4142	***NF1***	fw: 5’ CCCTGTTGTAAGTCCTATGG 3’ rev: 5’ CTTCATCAATTCCAGGCAGG 3’
4142	***EGFR***	fw: 5’ AATCCAACAAATGTGAACGG 3’ rev: 5’ AACTGAACCTGTGACTCACC 3’
4142	***TP53***	fw: 5’ CCTAAGAGCAATCAGTGAGG 3’ rev: 5’ TGCAGCTGTGGGTTGATTCC 3’
4142	***CIC***	fw: 5’ CAGCCTTCTCAAGGGGTCTGC 3’ rev: 5’ CTGCTCTCGCTGCTGCCACC 3’
4142	***ATRX***	fw: 5’ GCATGTGCTCACTATCTACC 3’ rev: 5’ TTCCGAGTTTCGAGCGATGG 3’
4153	***TP53***	fw: 5’ TGTTGTTGGGCAGTGCTAGG 3’ rev: 5’ GTGGTAATCTACTGGGACGG 3’

4153	***TP53***	fw: 5’ TGTCCCAGAATGCAAGAAGC 3’ rev: 5’ TTTCACCCATCTACAGTCCC 3’
4153	***ATRX***	fw: 5’ ACCATAGTCTACTGTACTGG 3’ rev: 5’ TAAATTTTCGTCAGGTCTGC 3’
4153	***ATRX***	fw: 5’ GCATGTGCTCACTATCTACC 3 rev: 5’ GAAGGCTCATCTTGCATTGG 3’
4153	***ACVR1***	fw: 5’ TGGATTCCATTCTGACAACC 3’ rev: 5’ AAGCCAGTTTGTCATTGTGG 3’
4155	***TP53***	fw: 5’ TGGAAGAAATCGGTAAGAGG 3’ rev: 5’ CTTGGGCCTGTGTTATCTCC 3’
4155	***ATRX***	fw: 5’ CTATGGAACATATTTGTACC 3’ rev: 5’ TCTCTTAGATCATTGTATGG 3’
4158	***PTEN***	fw: 5’ ACCACAGTTGCACAATATCC 3’ rev: 5’ AAATCTAGGGCCTCTTGTGC 3’
4158	***TP53***	fw: 5’ AGGCATAACTGCACCCTTGG 3’ rev: 5’ TGGGAGTAGATGGAGCCTGG 3’
4158	***ATRX***	fw: 5’ AGAACTGTGACTCATCCTGC 3’ rev: 5’ ATGTAATGAAACAGTTAAGG 3’
4160	***NF1***	fw: 5’ TTTCTCCTAGGTCAGCTGCC 3’ rev: 5’ CGTGAGGTGTGGCTCATTGG 3’
4160	***TP53***	fw: 5’ GCTTTCCAACCTAGGAAGGC 3’ rev: 5’ TTACTTCTCCCCCTCCTCTG 3’
4160	***ATRX***	fw: 5’ TTTACAGCATCCATCGCTCG 3’ rev: 5’ CCTGTTCTGGCTCTGTAACC 3’
4160	***PIK3R1***	fw: 5’ GGAAATGATCGATGTGCACG 3’ rev: 5’ GCAAGACATATACAAGCACC 3’
P-3879	***NF1***	fw: 5’ CAAATATATGTCTTCCACCC 3’ rev: 5’ TGAATGTGTTATAGTTGGGC 3’
P-3879	***CIC***	fw: 5’ CACCATGGTCACCAATGTGG 3’ rev: 5’ AGGTTAGTGACAGTGGCAGG 3’
P-3879	***ATRX***	fw: 5’ ATATGTTTACCTTTGGCAGC 3’ rev: 5’ TTCATATTAACCAGTAACCG 3’
R-77342	***NF1***	fw: 5’ TTTTCTCCTAGGTCAGCTGC 3’ rev: 5’ ATTCTAACGTGAGGTGTGGC 3’
R-77376	***ACVR1***	fw: 5’ AAGAATCGAAACAATCCACC 3’ rev: 5’ AGTGTATTGCAACAGTGACC 3’
R-78222	***TP53***	fw: 5’ GTTGCAAACCAGACCTCAGG 3’ rev: 5’ATGAGCGCTGCTCAGATAGC 3’

R-78610	***PIK3R1***	fw: 5’ GGAAGAAGACTTGAAGAAGC 3’ rev: 5’ TTTCATTGCCCAACCACTCG 3’
R-78653	***NF1***	fw: 5’ GCATGAGAAATCATTCTAGC 3’ rev: 5’ AACATTCAACACTGATACCC 3’
R-78653	***TP53***	fw: 5’ CAGGCATTGAAGTCTCATGG 3’ rev: 5’ CAATGGTTCACTGAAGACCC 3’
R-78653	***ATRX***	fw: 5’ CAAATTTCTTCTCGCTCAGG 3’ rev: 5’ AGAGCAAGCATCTCAAAACC 3’

A.8 Results of validation of the selected NGS novel mutations with Sanger sequencing

Illustrative tables representing each of NGS novel mutations selected with its major features (variant frequency and read depths) and the Sanger sequencing results – if the mutation is real (confirmed) or if it’s a false positive (WT). **I** represents the first 22 mutations analysed, which only 9% were confirmed and **II** represents the further 81 mutations selected, which 40,5% were validated.

I. Results of the first set of Sanger’s validations

DNA	Gene	Mutation	Alt var freq (%)	Read depth	Result
P4101	*BRAF*	C>C/T	20,84	403	WT
P4101	*CIC*	C>C/T	33,7	629	WT
P4101	*CIC*	C>C/T	36,24	367	WT
P4101	*CIC*	G>G/A	46,84	1106	WT
P4101	*CIC*	C>C/T	20,8	697	WT
R78610	***CIC***	**G>G/GC**	**96,19**	**105**	**CONFIRMED**
P4101	*EGFR*	TC>TC/T	22,49	498	WT
P4101	*FUBP1*	G>G/A	44,19	525	WT
P4101	*FUBP1*	G>G/A	43,38	521	WT
P4101	*FUBP1*	C>C/T	27,18	916	WT
P4101	*FUBP1*	C>C/T	28,95	950	WT
P4101	*FUBP1*	G>G/A	60,54	223	WT
P4099	*HIST1H3B*	C>C/T	24,08	1711	WT
P4101	*HIST1H3B*	C>C/T	33,62	580	WT

P4101	*HIST1H3B*	C>C/T	32,88	587	WT
P4101	*H3F3A*	C>C/T	25,9	2849	WT
P4099	*IDH1*	C>C/T	28,25	1521	WT
P4099	*IDH1*	G>G/A	46,57	3487	WT
P4102	*IDH1*	C>C/T	23,05	1388	WT
P4107	***PIK3R1***	**CTCAGTT> CTCAGTT/C**	**97,25**	**255**	**CONFIRMED**
P4107	*PIK3R1*	C>C/T	98,68	228	WT
P4101	*PPM1D*	C>C/T	45,51	690	WT

II. Results of the second set of Sanger's validations

DNA	Gene	Mutation	Alt var freq (%)	Read depth	Result
P4125	*ACVR 1*	G>G/A	34,9	1209	WT
P4153	***ACVR1***	**G>G/A**	**50,18**	**1706**	**CONFIRMED**
R77376	***ACVR1***	**C>C/T**	**39,19**	**694**	**CONFIRMED**
P4099	*ACVR1*	CG>C	43,1	928	WT
P4109	*ACVR1*	T>T/C	22,02	1694	WT
P3879	***ATRX***	**C>C/T**	**34,52**	**1008**	**CONFIRMED**
P4119	*ATRX*	T>T/A	43,88	4756	WT
P4121	*ATRX*	C>C/T	51,96	460	WT
P4121	*ATRX*	G>G/A	42,52	1129	WT
P4125	*ATRX*	G>A/A	99,85	682	WT
P4142	*ATRX*	T>T/A	26,12	1252	WT
P4153	***ATRX***	**G>G/A**	**76,62**	**1065**	**CONFIRMED**
P4153	***ATRX***	**T>T/G**	**92,37**	**2149**	**CONFIRMED**
P4155	*ATRX*	G>G/T	32,05	571	WT
P4158	***ATRX***	**C>C/G**	**35,2**	**2835**	**CONFIRMED**
P4160	*ATRX*	G>G/A	21,67	1117	WT
P4160	*ATRX*	G>G/A	22,6	1137	WT
P588	*ATRX*	TTCC>TTCC/T	92,85	3663	WT
R78653	***ATRX***	**TAA>TAA/T**	**90,79**	**3126**	**CONFIRMED**
P585	*ATRX*	TTCC>TTCC/T	3,06	686	WT
P748	***ATRX***	**G>G/A**	**72,79**	**2870**	**CONFIRMED**
P4101	*ATRX*	CT>CT/C	24,96	669	WT
P4102	*ATRX*	CG>CG/C	37,82	743	WT
P4111	*ATRX*	TG>TG/T	22,68	626	WT
P4121	*CIC*	C>T/T	100	165	WT
P4142	*CIC*	G>G/A	28,98	1042	WT
P4099	*CIC*	G>G/C	24,82	1410	WT

P3879	*CIC*	G>G/A	46,13	388	WT
P4129	*EGFR*	C>C/T	21,58	709	WT
P4142	*EGFR*	C>C/T	25,63	1662	WT
P585	***EGFR***	**C>C/T**	**49,24**	**3109**	**CONFIRMED**
P4120	*ERBB2*	G>A/A	100	139	WT
P4125	*H3F3A*	G>G/A	66,22	1350	WT
P4118	*NF1*	T>C/C	100	584	WT
P4121	*NF1*	G>G/A	59,92	781	WT
P4121	***NF1***	**C>C/T**	**40,68**	**799**	**CONFIRMED**
P4121	*NF1*	ACT>ACT/A	52,94	850	WT
P4121	*NF1*	AG>AG/A	34,51	988	WT
P4125	*NF1*	G>A/A	99,12	455	NO PCR PRODUCT
P4129	*NF1*	G>G/T	21,19	670	WT
P4142	*NF1*	C>C/T	32,2	531	WT
P4160	*NF1*	G>G/T	35,36	577	WT
P588	***NF1***	**AC>AC/A**	**55,16**	**7408**	**CONFIRMED**
R78653	***NF1***	**CACTT>CACTT/C**	**77,35**	**1289**	**CONFIRMED**
P4107	*NF1*	C>T/T	100,00	512	WT
P4109	*NF1*	G>G/T	34,78	1225	WT
P3879	***NF1***	**A>A/AC**	**24,3**	**7753**	**CONFIRMED**
R77342	*NF1*	G>G/T	31,21	676	WT
P4099	*PIK3CA*	C>C/T	29,35	1550	WT
P4102	*PIK3CA*	A>A/T	30,61	735	WT
P4160	*PIK3R1*	GA>GA/G	25,38	1249	WT
R78610	***PIK3R1***	**AGAC>AGAC/A**	**46,76**	**1929**	**CONFIRMED**
P4099	*PIK3R1*	G>G/A	30,37	1936	WT
P4102	*PPM1D*	AG>AG/A	26,04	1102	WT
P4125	*PTEN*	C>C/T	45,12	1507	NO PCR PRODUCT
P4158	*PTEN*	G>G/A	33,18	434	WT
P585	***PTEN***	**C>C/T**	**82,27**	**908**	**CONFIRMED**
P748	***PTEN***	**C>C/T**	**62,36**	**2309**	**CONFIRMED**
P4110	***PTEN***	**CCAGT>CCAGT/C**	**40,75**	**1460**	**CONFIRMED**
P4118	*TP53*	AG>AG/A	96,84	412	WT
P4119	*TP53*	C>C/T	92,85	1049	WT
P4120	*TP53*	C>C/T	27,84	862	WT
P4121	***TP53***	**C>T/T**	**100**	**1737**	**CONFIRMED**
P4125	*TP53*	T>T/A	97,76	491	WT
P4129	***TP53***	**T>T/C**	**72,51**	**2899**	**CONFIRMED**

P4142	***TP53***	**G>G/C**	**27,97**	**6195**	**CONFIRMED**
P4153	***TP53***	**G>G/A**	**48,56**	**1979**	**CONFIRMED**
P4153	***TP53***	**G>G/A**	**47,57**	**3786**	**CONFIRMED**
P4155	***TP53***	**C>C/T**	**39,83**	**4115**	**CONFIRMED**
P4158	***TP53***	**G>G/A**	**50,68**	**1178**	**CONFIRMED**
P4160	***TP53***	**G>G/A**	**98,69**	**153**	**CONFIRMED**
P588	***TP53***	**C>C/T**	**57,11**	**11955**	**CONFIRMED**
R78222	***TP53***	**G>G/A**	**42,4**	**1988**	**CONFIRMED**
R78653	***TP53***	**A>A/C**	**54,98**	**2008**	**CONFIRMED**
P585	***TP53***	**C>C/T**	**74,04**	**11599**	**CONFIRMED**
P748	***TP53***	**C>C/T**	**66,51**	**7095**	**CONFIRMED**
P4102	*TP53*	T>T/C	32,00	1847	WT
P4107	*TP53*	C>T/T	99,42	859	WT
P4109	***TP53***	**A>A/C**	**76,2**	**6526**	**CONFIRMED**
P4110	***TP53***	**A>A/AG**	**30,71**	**1234**	**CONFIRMED**
P4099	*TPP53*	G>G/A	32,21	2136	WT